



**UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
DEPARTAMENTO DE BIOLOGIA**

TATIANE EMILIANO SANTOS

**ANÁLISE DO PADRÃO DE USO DE CÓDONS SINÔNIMOS EM
PAPILOMAVÍRUS HUMANO**

SÃO CRISTÓVÃO

2015/2

TATIANE EMILIANO SANTOS

**ANÁLISE DO PADRÃO DE USO DE CÓDONS SINÔNIMOS EM
PAPILOMAVÍRUS HUMANO**

Monografia apresentada ao Curso de Bacharelado em Ciências Biológicas do Departamento de Biologia da Universidade Federal de Sergipe como requisito parcial para obtenção do título de Bacharel em Ciências Biológicas.

Orientador: Prof. Dr. Marcus Vinicius de Aragão Batista

SÃO CRISTÓVÃO

2015/2

RESUMO

O câncer de colo de útero é o segundo tipo de câncer mais comum em mulheres vivendo em regiões em desenvolvimento. A principal causa de câncer de colo de útero é a infecção por papilomavírus humano (HPV). O objetivo deste trabalho é investigar o papel do uso de códons sinônimos na evolução do papilomavírus humano e seu hospedeiro, utilizando as diferentes variantes virais, visando identificar possíveis padrões adaptativos associados à sua patogenicidade. Foram calculados para este estudo a utilização relativa de códons sinônimos (RSCU), número efetivo de códons (ENC), análise de componentes principais (PCA), análise de correlação e análise de agrupamentos. Foi verificado um viés na utilização dos códons em HPV, evidenciando que a frequência de códons sinônimos não é a mesma para o mesmo aminoácido. Os diferentes gêneros de HPV mostram uma utilização de códons semelhantes entre si, porém não coincidente com a do hospedeiro, corroborando com resultados de pesquisas de outros autores. O gênero *Alphapapillomavirus* apresentou uso de códons diferente dos demais gêneros. Esse gênero possui os tipos de HPV de alto risco, e essa utilização diferenciada de códons pode estar relacionada com a sua adaptação ao hospedeiro, permitindo aos vírus de *Alphapapillomavirus* graus maiores de patogenicidade. Outros estudos ainda precisam ser realizados para que se determine as possíveis relações entre os padrões de uso de códons em HPV, sua diversificação, e a associação com os diferentes níveis de patogenicidade dos diferentes tipos e variantes virais.

Palavras-chave: papilomavírus humano (HPV), códons sinônimos, viés de uso de códons

SUMÁRIO

1. INTRODUÇÃO	4
1.1 Câncer de Colo de Útero.....	4
1.2 Propriedades Gerais do HPV	4
1.3 Ciclo de Infecção do HPV	5
1.4 Evolução do Papilomavírus	6
1.5 Utilização de Códon Sinônimos.....	10
2. OBJETIVOS	13
2.1 Objetivo Geral.....	13
2.2 Objetivos Específicos	13
3. METODOLOGIA.....	14
3.1 Obtenção das sequências virais e da frequência de uso de códon do hospedeiro	14
3.2 Utilização relativa de códon sinônimos (RSCU)	15
3.3 Número efetivo de códon (ENC)	15
3.4 Análise de componentes principais (PCA)	15
3.5 Análise de correlação e análise de agrupamentos (cluster)	15
3.6 Análise estatística	15
4. RESULTADOS	16
4.1 Viés de utilização de códon e utilização de códon sinônimos em HPV	16
4.2 Propriedades de composição em HPV	19
4.3 Análise de correlação	20
4.4 Análise de componentes principais (PCA)	22
4.5 Análise de agrupamentos	26
5. DISCUSSÃO	33
6. CONCLUSÃO.....	36
7. REFERÊNCIAS BIBLIOGRÁFICAS	37
Apêndice A	40

1. INTRODUÇÃO

1.1 Câncer de Colo de Útero

O câncer de colo de útero é o segundo tipo de câncer mais comum em mulheres vivendo em regiões em desenvolvimento. Em 2012, foram registrados 528.000 novos casos de câncer de colo de útero e aproximadamente 266.000 mulheres morreram vítimas desse tipo de câncer no mundo, sendo que mais de 85% dos casos ocorreu em países de baixa a média renda (WHO, 2014). No Brasil, estima-se que 18.503 mulheres são diagnosticadas e 8.414 morrem a cada ano vítimas de câncer de colo de útero (INFORMATION CENTRE ON HPV, 2014). A principal causa do câncer de colo de útero é a infecção (livre de sintomas, persistente ou crônica) pelo Papilomavírus Humano (HPV), que é a infecção sexualmente transmissível mais comum no mundo (WHO, 2014).

1.2 Propriedades Gerais do HPV

O HPV é um vírus da família Papillomaviridae, não envelopado, com diâmetro de aproximadamente 55 nm e possui DNA circular de fita dupla, com aproximadamente oito sequências de leitura abertas (ORFs ou genes), envolto num capsídeo proteico composto por duas proteínas: L1 e L2. Essas duas proteínas estruturais são transcritas pela região tardia (L) dos genes e as proteínas E1, E2 e E4 (regulatórias), que modulam a transcrição e replicação e E5, E6 e E7 (oncoproteínas), que modulam o processo de transformação são transcritas pela região precoce dos genes (SANTOS, ROMANOS & WIGG, 2008; IARC, 2005; SANGAR et al., 2014; DE VILLIERS et al., 2004), além também de uma região não codificante, chamada de região longa de controle regulatório (LCR) (BERNARD, CALLEJA-MACIAS & DUNN, 2006).

A família Papillomaviridae é dividida em 29 gêneros, que infectam a maioria dos mamíferos e aves (DE VILLIERS et al., 2004). Os papilomavírus (PVs) encontrados em humanos são divididos em cinco gêneros: Alpha, Beta, Gamma, Mu e Nupapillomavirus. O gênero Alphapapillomavirus causa lesões cutâneas e mucosas, e inclui tipos de alto risco (associados ao câncer de colo de útero) e baixo risco (associados a lesões benignas). Os gêneros Beta, Gamma, Mu e Nupapillomavirus causam, principalmente, lesões benignas (DE VILLIERS et al., 2004; BERNARD et al., 2010; DOOBAR et al., 2012).

O gene L1 tem sido utilizado para identificação de tipos, subtipos e variantes, por ser o gene mais conservado. Um tipo é reconhecido quando a sequência de DNA do gene L1 difere mais de 10% dos demais tipos. O termo espécie é usado para designar um conjunto de tipos que apresentam identidade de 60-70% nesse gene. Subtipos são definidos quando, dentro de um mesmo tipo, existe diferença entre 2-10% e variantes quando, dentro de um mesmo subtipo, há divergência de menos de 2% (DE VILLIERS et al., 2004; BERNARD, CALLEJA-MACIAS & DUNN, 2006).

Foram completamente sequenciados mais de 200 tipos de HPV, que causam diferentes lesões em diferentes locais do corpo (DOOBAR et al., 2012; WHO, 2014; IARC, 2005), porém o órgão genital feminino é o mais atingido, principalmente o colo do útero (BICHO, 2013). Doze tipos de HPV (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 e 59) foram definidos como tipos de alto risco, por serem causadores de câncer (DOORBAR et al., 2012). Porém, mais oito tipos foram incluídos nesta lista (26, 53, 66, 67, 68, 70, 73 e 82), sendo encontrados em casos raros, totalizando 20 tipos de HPV carcinogênicos (ARBYN et al., 2014). Os tipos 16 e 18 são responsáveis por 70% dos casos de câncer de colo de útero no mundo (WHO, 2014) e 68,5% no Brasil (INFORMATION CENTRE ON HPV, 2014).

1.3 Ciclo de Infecção do HPV

A infecção causada pelo HPV é exclusivamente intraepitelial (BICHO, 2013) e sua replicação ocorre no núcleo celular do hospedeiro, utilizando a maquinaria celular para seu propósito e adaptando-se ao tecido natural do hospedeiro (SANGAR et al., 2014; SANTOS, ROMANOS & WIGG, 2008). A população viral pode permanecer por longo tempo no hospedeiro sem causar danos, indicando que ocorre um balanço entre a replicação viral e a tolerância imunológica na interação vírus-hospedeiro (DOORBAR et al., 2012).

O ciclo de vida se inicia pela infecção das células epiteliais (queratinócitos) basais, que frequentemente levam a microlesões, que podem ser visíveis ou não. Porém alguns tipos de HPV necessitam da presença pré-existente das microlesões para infectar a célula (DOORBAR et al., 2012; DE VILLIERS et al., 2004; IARC, 2005). O reservatório da infecção é formado pelas células basais infectadas, onde o genoma viral é mantido, dependente das proteínas E1, E2, E6 e em alguns casos E7, em um baixo número de cópias. Os tipos de alto risco possuem a habilidade de persistirem por anos, por razões ainda não conhecidas, e conduzir a proliferação celular nas camadas basais e parabasais em alguns locais infectados. Estas células se dividem, produzindo células filhas, que são empurradas em

direção à superfície epitelial. Nas camadas médias, as proteínas necessárias para a amplificação do genoma se tornam elevadas, o que permite que a amplificação do genoma ocorra. Estas células expressam a proteína E4 viral e estão geralmente na fase S ou G2 do ciclo celular. Nas camadas epiteliais superiores, as células saem do ciclo celular, e em um subconjunto de células E4-positivas, as proteínas L1 e L2 do vírus são transcritas, permitindo a disposição dos genomas virais amplificados (SANGAR et al., 2014; DOORBAR et al., 2012).

Antes da reinfecção, o vírus necessita ser liberado no meio ambiente através da superfície epitelial e sobreviver extra-celularmente. A liberação pode ocorrer dentro de uma escama cornificada, o que melhora a sobrevivência extra-celular do vírus. Isso pode causar o comprometimento da detecção imunológica do vírus e a consequente limitação da apresentação de epítomos virais para o sistema imunitário nas camadas epiteliais inferiores (SANGAR et al., 2014).

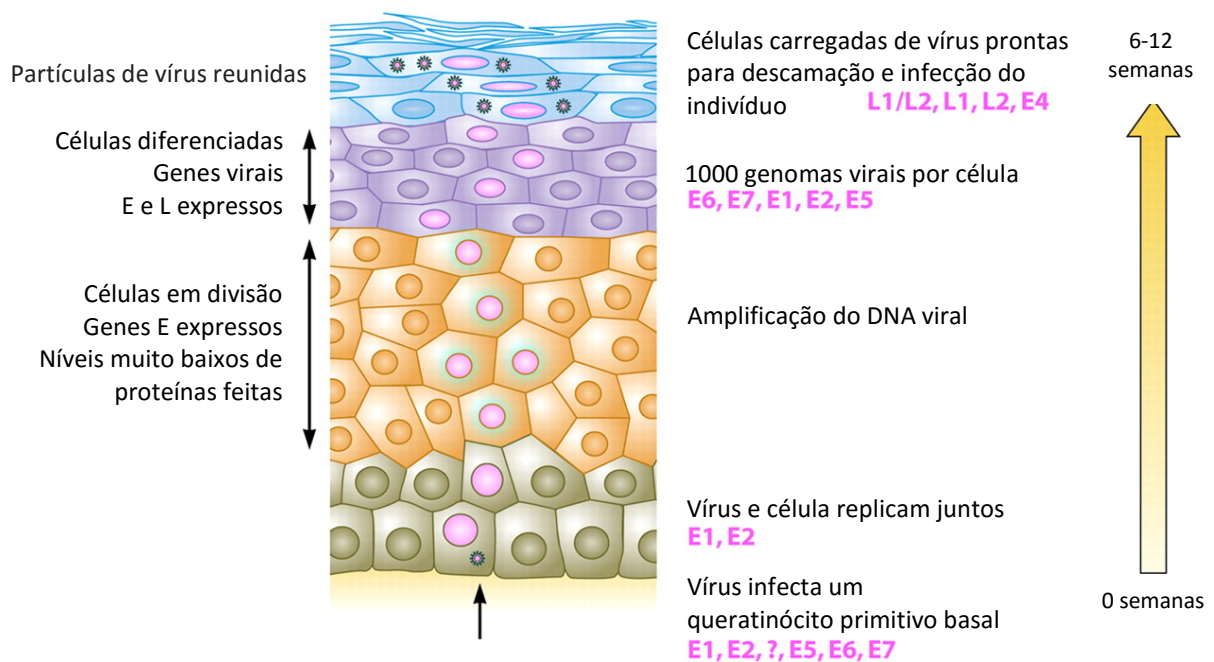


Fig. 1 – Ciclo de vida de papilomavírus.
Fonte: modificado de STANLEY, 2012

1.4 Evolução do Papilomavírus

Uma linhagem de PVs, provavelmente com tropismo mucoso, deu origem ao supertaxon compreendendo Alpha e Omicronpapillomavirus, que inclui vírus que infectam primatas, carnívoros, suínos e cetáceos. Uma segunda linhagem deu origem a um grande clado de PVs, principalmente com tropismo cutâneo, incluindo Beta e Xipapillomavirus, que

infectam primatas, roedores, carnívoros, ruminantes, ouriços e marsupiais. Uma terceira linhagem deu origem ao supertaxon de Delta e Zetapapillomavirus, que infectam perissodáctilos e ruminantes. Uma quarta linhagem deu origem ao supertaxon de Kappa e Lambdapapillomavirus, que infectam primatas, roedores, lagomorfos e carnívoros (BRAVO, DE SANJOSÉ & GOTTSCHLING, 2010).

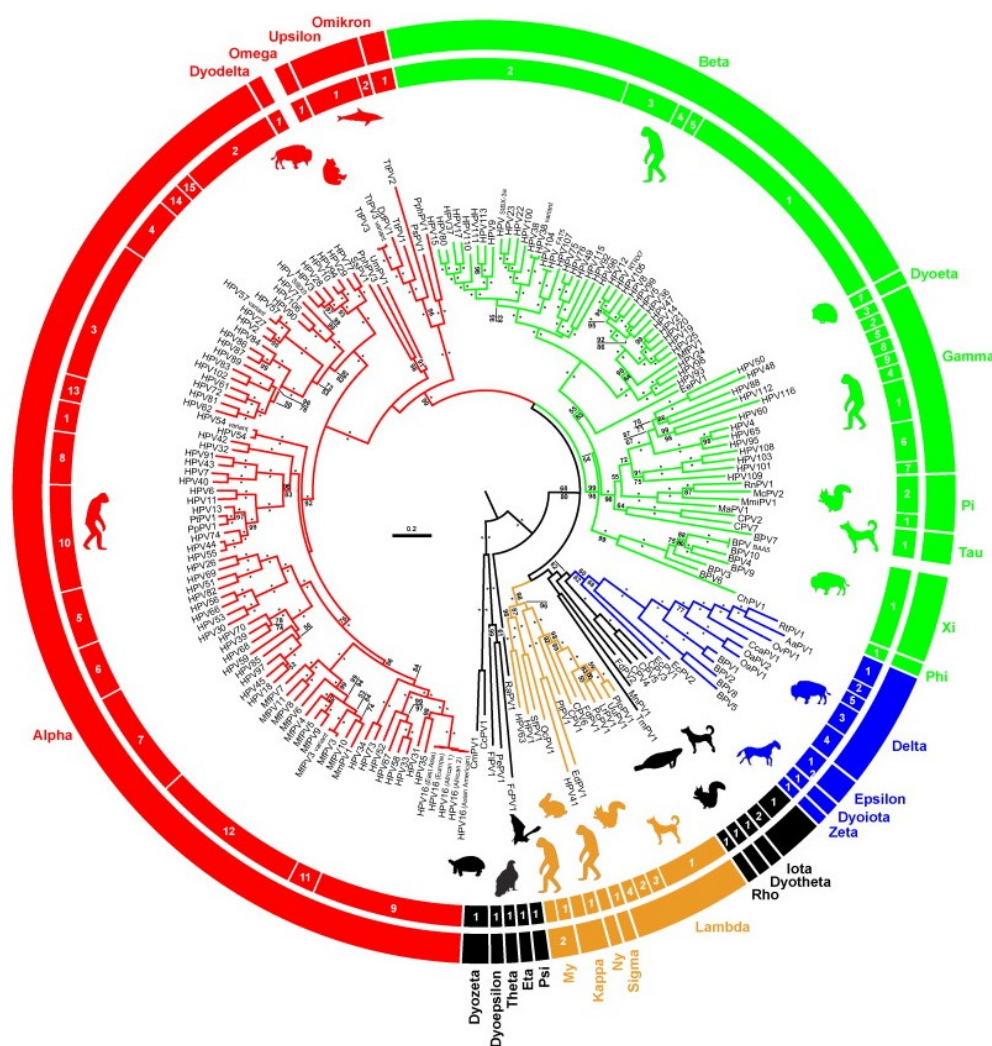


Fig. 2 – Árvore filogenética de papilomavírus
Fonte: BRAVO, 2010

É possível que muitas linhagens de PVs não-humanos foram extintas, mas é difícil explicar esse padrão de extinção dado o grau de diversificação de PVs observado entre os seres humanos (SHAH, DOORBAR & GOLDSTEIN, 2010).

Apesar de mamíferos serem infectados por diversos tipos de PVs, é pressuposto que os vírus são transmitidos somente na vertical (para descendentes) e são hospedeiro-específicos. Portanto, patógenos de humanos têm origem em patógenos de primata, o que

significa que o último ancestral comum de *Homo sapiens* foi infectado (IARC, 2005; GOTTSCHLING et al., 2011).

Papilomavírus pode ter coevoluído estreitamente com seus hospedeiros mamíferos (BRAVO, DE SANJOSÉ & GOTTSCHLING, 2010; GOTTSCHLING et al., 2011). Porém, os PVs que infectam as mesmas espécies muitas vezes não são monofiléticos, mas parafiléticos ou polifiléticos, como o HPV, principalmente Alphapapillomavirus, que tem parentes próximos em diferentes espécies hospedeiras, não-humanas (IARC, 2005; BRAVO, DE SANJOSÉ & GOTTSCHLING, 2010; GOTTSCHLING et al., 2011). Além de que alguns grupos são congruentes com o táxon correspondente dos mamíferos infectados por eles, porém muitos não são agrupados de acordo com a filogenia de seus hospedeiros (IARC, 2005; GOTTSCHLING et al., 2007).

Entretanto, coevolução não é um mecanismo suficiente para explicar por si só este padrão de evolução, apesar de explicar quase um terço de todos os eventos para conciliar as histórias evolutivas de PVs e seus hospedeiros. Outros mecanismos também contribuíram para a diversificação de PVs, como radiações adaptativas (duplicações) e ocupação de novos nichos ecológicos, assim como recombinação, apesar de o potencial de recombinação de PVs ser considerado baixo, além de convergência de genomas assimétricos impulsionada pela seleção intensa e utilização de códons sinônimos (GOTTSCHLING et al., 2007; BRAVO, DE SANJOSÉ & GOTTSCHLING, 2010; GOTTSCHLING et al., 2011; BRAVO & FÉLEZ-SÁNCHEZ, 2015).

A transferência de hospedeiro acontece quando ou a totalidade ou parte de um genoma viral adaptado para uma espécie hospedeira infecta outra, resultando em uma linhagem estável no novo hospedeiro (SHAH, DOORBAR & GOLDSTEIN, 2010). A pele de mamíferos pode ter fornecido novos habitats para os vírus, sob a forma de novos tipos de células ou de novos microambientes histológicos e novos nichos ecológicos (BRAVO, DE SANJOSÉ & GOTTSCHLING, 2010). A estreita relação evolutiva pode facilitar infecções virais em um novo hospedeiro, com taxa de reprodução baixa. Porém, a infecção viral pode também ocorrer em hospedeiros distantemente relacionados por mudanças dramáticas no comportamento do hospedeiro (GOTTSCHLING et al., 2011).

A especificidade rigorosa de hospedeiro pode agir como uma barreira que impede o contato íntimo entre diferentes vírus, mas vários PVs infectam uma variedade de hospedeiros filogeneticamente distantes (infecções heterólogas). Essa mudança reduz as pressões evolutivas sobre os PVs presentes no hospedeiro original e pode permitir mudanças no tempo de infecção e virulência (BRAVO, DE SANJOSÉ & GOTTSCHLING, 2010).

Endotermia e um estado imunológico baixo podem permitir que o vírus rompa essa barreira entre espécies (GOTTSCHLING et al., 2007). As alterações genéticas em sequências virais recuperadas de cavalos sugerem um processo de adaptação ao ambiente proporcionado pelo novo hospedeiro. A domesticação de animais, assim como de seres humanos, pode ter, também, propiciado novos nichos ecológicos para as novas linhagens de PVs (BRAVO, DE SANJOSÉ & GOTTSCHLING, 2010).

O cenário evolutivo que se adapta melhor à descrição atual da diversidade genética de PVs é uma série de eventos de duplicação basais seguidos de limitada coevolução (BRAVO & FÉLEZ-SÁNCHEZ, 2015). A duplicação (radiação adaptativa) é explicada quando ocorre uma divergência prévia, ou seja, uma única linhagem viral irradia em múltiplas linhagens virais dentro de um único ancestral hospedeiro. Quando ocorre especiação posterior do hospedeiro, as diversas linhagens segregam diferencialmente, podendo ocorrer a infecção com várias linhagens ou a extinção de algumas dessas linhagens (SHAH, DOORBAR & GOLDSTEIN, 2010).

A taxa lenta de evolução do PVs pode prejudicar uma adaptação rápida a um novo hospedeiro (SHAH, DOORBAR & GOLDSTEIN, 2010; BRAVO & FÉLEZ-SÁNCHEZ, 2015), por isso, a geração de diversidade dentro do hospedeiro ocorre durante as infecções crônicas, pois, com a baixa taxa de evolução, o vírus necessita de muito tempo para diversificar. A replicação celular descontrolada de longa duração em infecções crônicas, eventualmente, conduz a uma acumulação de mutações e, em alguns casos, à instabilidade genômica associada à integração do genoma viral ao genoma celular (BRAVO & FÉLEZ-SÁNCHEZ, 2015).

Além disso, vários eventos de recombinação entre genes podem ter ocorrido, inclusive eventos de recombinação intra-gene, assim como recombinações diversas entre as linhagens virais ancestrais (SHAH, DOORBAR & GOLDSTEIN, 2010). Evidências moleculares sugerem que a recombinação é central para a replicação do genoma de PVs, e a filogenia e a genômica comparativa identificam vários eventos de recombinação ao longo da história. A replicação requer recombinação homóloga, que pode permitir eventos raros de recombinação não homóloga. Alphapapillomavirus clinicamente importantes têm sido submetidos a eventos de recombinação entre as regiões iniciais e as regiões finais do genoma, como consequência, as relações filogenéticas entre Alphapapillomavirus diferem quando inferidas com base nos genes iniciais ou nos genes finais (BRAVO & FÉLEZ-SÁNCHEZ, 2015).

Além da recombinação entre táxons da mesma família de vírus, pode ocorrer recombinação entre famílias próximas, por infecções cruzadas (GOTTSCHLING et al., 2011; BRAVO & FÉLEZ-SÁNCHEZ, 2015). Polyomaviridae também é um pequeno vírus de DNA circular, e a convergência de mecanismos semelhantes de replicação pode permitir eventos raros de recombinação não homóloga entre os vírus dessa família e os da família Papillomaviridae (BRAVO & FÉLEZ-SÁNCHEZ, 2015).

Outro mecanismo também responsável pela evolução de PVs é a utilização de códons sinônimos. Como o vírus utiliza a maquinaria celular para replicação, é importante que seus genes utilizem códons que possam ser traduzidos pelos tRNAs produzidos pela célula (ZHAO, LIU & FRAZER, 2003; ZHAO & CHEN, 2011). Porém, a maior parte dos códons utilizados pelo parasita não são os mesmos utilizados pelo hospedeiro, isto é chamado de viés na utilização de códons (ZHAO & CHEN, 2011; VAN DOORSLAER, 2013; BRAVO & FÉLEZ-SÁNCHEZ, 2015).

1.5 Utilização de Códons Sinônimos

O código genético é degenerado, pois um mesmo aminoácido pode ser codificado por códons diferentes (ALBERTS et al., 1999). As trincas de bases que codificam um mesmo aminoácido são chamados de códons sinônimos. Geralmente, códons de um mesmo aminoácido podem não ocorrer com a mesma frequência, indicando que existe utilização preferencial de códons (MATTIOLI, 2008).

U		C		A		G			U		C		A		G			
U	UUU	Phe	UCU	Ser (S)	UAU	Tyr	UGU	Cys	A	AUU	Ile (I)	ACU	Thr (T)	AAU	Asn (N)	AGU	Ser (S)	U
	UUC	(F)	UAC		(Y)	UGC	(C)	AUC		AAC		AGC		(S)	C			
	UUA	Leu	UCA	UAA	Stop	UGA	Stop	AUA		ACA	AAA	Lys (K)	AGA	Arg (R)	A			
	UUG	(L)	UCG	UAG	Stop	UGG	Trp (W)	AUG		Met (M)	AAG	AGG	(R)	G				
C	CUU	Leu (L)	CCU	Pro (P)	CAU	His (H)	CGU	Arg (R)	G	GUU	Val (V)	GCU	Ala (A)	GAU	Asp (D)	GGU	Gly (G)	U
	CUC		CAC		CGC	GUC	GCC			GAC		GGC		C				
	CUA		CAA	Gln (Q)	CGA	GUA	GCA			GAA	Glu (E)	GGA	A					
	CUG		CCG	CAG	CGG	GUG	GCG			GAG	GGG	G						

= Códons de parada

= Códon de início

© 2010 Pearson Education, Inc.

Fig. 3 – Código genético padrão

Fonte: Modificado de RUSSEL, 2010. Disponível em: https://www.mun.ca/biology/scarr/iGen3_06-07.html

Seria esperado que a taxa evolutiva do hospedeiro fosse próxima à dos parasitas, porém as preferências de utilização de códons em HPV não coincidem com as do genoma humano (BRAVO & MÜLLER, 2005; BRAVO & FÉLEZ-SÁNCHEZ, 2015). Esse desvio de utilização de códons não é aleatório (BRAVO & MÜLLER, 2005). A utilização diferencial de códons sinônimos poderia ajudar a determinar diferenças na expressão temporal. As proteínas E1 e E2 são produzidas no início do ciclo de vida do vírus. L1 e L2 são produzidas em células terminalmente diferenciadas (CLADEL, BERTOTTO & CHRISTENSEN, 2010). Além disso, a distribuição de códons sinônimos, ao longo da evolução, não é equitativa, alguns códons são mais utilizados que outros, tanto entre gêneros ou espécies diferentes quanto entre genes diferentes (BRAVO & MÜLLER, 2005; CLADEL, BERTOTTO & CHRISTENSEN, 2010; ZHAO & CHEN, 2011).

Geralmente, um tRNA maduro mostra um anticódon quimicamente modificado, e a interação códon-anticódon baseia-se, muitas vezes, em um pareamento não regular de bases nitrogenadas modificadas, na primeira posição do anticódon e na terceira posição do códon. Isso ocorre para que haja um padrão econômico de utilização de tRNAs (Fig. 4). Por exemplo, anticódons iniciados em A podem ter essa base nitrogenada trocada para inosina (I), que é capaz de interagir com as pirimidinas T e C, e também pode reconhecer a purina A (BRAVO & MÜLLER, 2005).

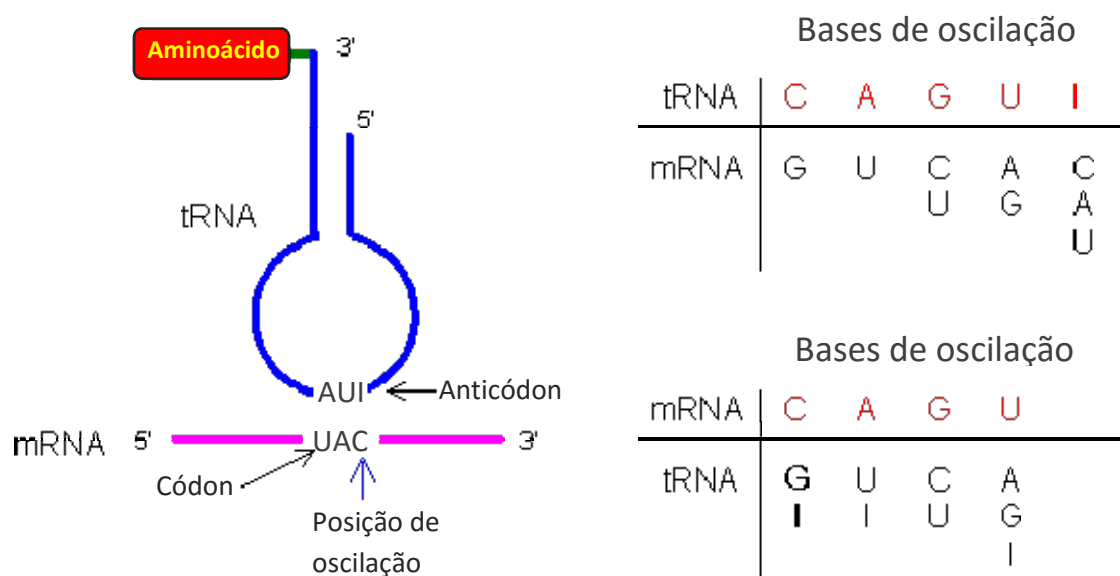


Fig. 4 - Emparelhamento entre o anticódon do tRNA e o códon do mRNA. A Fig. da esquerda define a posição de oscilação em que o emparelhamento de bases não obedece à regra padrão. As tabelas da direita mostram todos os emparelhamentos de base possíveis na posição de oscilação.

Fonte: modificado de <http://www.web-books.com/MoBio/Free/Ch5C4.htm>

As células humanas possuem um alto conteúdo GC, enquanto o genoma de HPV mostra um alto conteúdo AT, principalmente na terceira posição (ZHAO, LIU & FRAZER, 2003; BRAVO & MÜLLER, 2005). O viés na utilização dos códons está positivamente correlacionado com o conteúdo A+T na terceira posição para a maioria dos genes (ZHAO, LIU & FRAZER, 2003).

Todos esses dados apontam para um baixo nível de adaptação dos genes de HPV às preferências de utilização de códons de seus hospedeiros (CLADEL, BERTOTTO & CHRISTENSEN, 2010). Entretanto, Cladel et al. (2010) utilizaram apenas quatro genes de 72 tipos de duas espécies do vírus e Zhao et al. (2003) e Félez-Sánchez et al. (2015) utilizaram apenas tipos de HPV nos seus trabalhos, não incluindo as variantes virais. Como já foi relatado, variantes virais de um mesmo tipo de HPV podem apresentar diferentes graus de patogenicidade, é importante avaliarmos se há algum viés no uso de códons sinônimos destas variantes que possam auxiliar no ciclo de infecção.

Além disso, a compreensão dos mecanismos de condução da mudança evolutiva no genoma viral se torna importante para entendermos como os vírus se diversificam, e assim criar estratégias de intervenção eficazes, contribuindo com o desenvolvimento de métodos de tratamento contra o câncer de colo de útero e outras doenças causadas pelo HPV, como o desenvolvimento de vacinas eficientes, além de desenvolver formas de diagnóstico e tratamentos eficazes.

Portanto, este trabalho se justifica por apresentar um estudo com maior representatividade de amostragem, incluindo genomas completos de tipos e variantes de HPV, visando encontrar indícios de coevolução com o hospedeiro através da utilização de códons sinônimos.

2. OBJETIVOS

2.1 Objetivo geral

Investigar o papel do uso de códons sinônimos na evolução do Papilomavírus humano e seu hospedeiro, utilizando as diferentes variantes virais, visando identificar possíveis padrões adaptativos associados à sua patogenicidade.

2.2 Objetivos específicos

- Analisar a composição nucleotídica dos genes estudados.
- Identificar o possível viés de utilização de códons sinônimos dos diferentes tipos e variantes de HPV.
- Verificar o nível de similaridade e divergência no uso de códons entre os diferentes grupos de HPV.
- Averiguar a existência de convergência entre os genomas dos HPVs e o humano.

3. METODOLOGIA

3.1 Obtenção das sequências virais e da frequência de uso de códons do hospedeiro

As sequências virais (Apêndice A) foram obtidas a partir do banco de dados público *PaVE* (*Papillomavirus Episteme*) do *GenBank*, disponível em <http://pave.niaid.nih.gov>. A frequência de códons e o conteúdo GC de *Homo sapiens* foram obtidos a partir do banco de dados de uso de códons *Kazusa*, disponível em <http://www.kazusa.or.jp/codon>. Todas as sequências obtidas para o gene E5 pertencem a tipos do gênero *Alphapapillomavirus*.

As sequências foram alinhadas no software MEGA7 e nove sequências foram excluídas da análise por possuírem grande divergência em relação às demais (Tabela 1).

Tabela 1 – Sequências excluídas

Tipo	Variante	Gênero	Número de acesso <i>GenBank</i>
HPV 34	-	<i>Alpha</i>	X74476.1
HPV 52	IN141070	<i>Alpha</i>	HQ537743.1
HPV 52	Qv15145	<i>Alpha</i>	HQ537739.1
HPV 54	-	<i>Alpha</i>	U37488.1
HPV 54	AE9	<i>Alpha</i>	AF436129.1
HPV 66	-	<i>Alpha</i>	U31794.1
HPV 67	Qv25738	<i>Alpha</i>	HQ537783.1
HPV 73	QV29446	<i>Alpha</i>	KF436836.1
HPV 82	IS39/AE2	<i>Alpha</i>	AF293961.1

3.2 Utilização relativa de códons sinônimos (RSCU)

Utilização relativa de códons sinônimos (RSCU) é uma medida simples para avaliar o viés na utilização de códons sinônimos, evitando a influência de confusão da composição de aminoácidos em diferentes genes, a qual é calculada pela razão da frequência observada e a frequência esperada de um códon, se a utilização do códon for uniforme, dentro de um conjunto de códons sinônimos. Quando o valor RSCU é maior que 1,0 indica que este desvio de códons tem alta frequência, se o valor for menor que 1,0 tem baixa frequência. Quando o valor de RSCU é igual a 1,0 significa que este códon tem uma seleção aleatória de tradução.

O valor RSCU foi usado para comparar a preferência de utilização de códons sinônimos nos diferentes gêneros de HPV com a preferência de utilização de códons do hospedeiro (*Homo sapiens*). Todos os valores foram assentados em tabelas.

3.3 Número efetivo de códons (ENC)

O número efetivo de códons (ENC) é um dos índices mais úteis, que quantificam o viés absoluto na utilização de códons sinônimos. Os valores ENC variam de 20 (em que apenas um códon sinônimo é escolhido pelo aminoácido correspondente) a 61 (quando todos os códons sinônimos são igualmente utilizados). Quanto maior for o valor ENC num gene, menor é a preferência de códons (WRIGH, 1990 apud XU et al., 2015). Neste estudo, foi utilizada esta abordagem para avaliar o grau do viés de utilização de códons em HPV.

3.4 Análise de componentes principais (PCA)

A análise de componentes principais (PCA) é um método estatístico multivariado comumente utilizado (WOLD et al., 1987 apud XU et al., 2015), que é utilizado para analisar as principais tendências do padrão de uso de códons. Os valores RSCU de cada gene foram representados como vetores dimensionais, excluindo AUG (Met), UGG (Trp) e os três códons de parada, de onde foram extraídos os componentes principais. Os resultados foram plotados em gráficos para investigação visual por gênero.

3.5 Análise de correlação e análise de agrupamentos (cluster)

Análise de correlação de Spearman foi utilizada para questionar a relação entre cada composição geral de nucleotídeos (T%, A%, C%, G%), a composição dos nucleotídeos na terceira posição (T₃%, A₃%, C₃%, G₃%), GC₁₂%, CG₃%, os componentes principais (PCA), hidrofobicidade (GRAVY) e aromaticidade dos aminoácidos. A análise de agrupamentos foi realizada para investigar a relação da variação do uso de códon nos diferentes genótipos de HPV.

3.6 Análise estatística

O software CodonW foi utilizado para calcular os índices de utilização de códons (RSCU e ENC), a composição de nucleotídeos (T%, A%, C%, G%, T₃%, A₃%, C₃%, G₃%, CG₃%). Todas as análises estatísticas foram realizadas pelo software estatístico SPSS 24.0, incluindo PCA, análise de correlação, bem como a análise de agrupamentos.

4. RESULTADOS

4.1 Viés de utilização de códons e utilização de códons sinônimos em HPV

Analisando os valores de RSCU (Tabela 2), pode-se notar, numa média geral de todos os genes, que todos os códons preferenciais utilizam A/T na terceira posição (numa proporção de 9:9). Numa análise gene a gene, os genes E1, E6, L1 e L2 utilizam somente A/T na terceira posição (sendo A mais utilizado que T nos genes precoces, na proporção de 11:7, e T mais utilizado que A nos genes tardios, nas proporções de 13:5 e 12:6, respectivamente), enquanto E2, E5 e E7 têm preferência por A/T na terceira posição (proporções de 16:2, 16:2 e 15:3, respectivamente), porém não exclusivamente (proporção de G/C de 1:1, 1:1 e 3:0). Na análise por gênero nota-se também a preferência por A/T na terceira posição, com preferência semelhante entre os gêneros, apenas com algumas dissimilaridades (Tabela 3).

Comparando a utilização de códons por gênero com a utilização de códons do hospedeiro (*Homo sapiens*) foi verificado que os códons utilizados pelo hospedeiro não são os mesmos utilizados pelo vírus, exceto para alguns aminoácidos no gênero *Nupapillomavirus* (Tabela 3). O conteúdo AT (média geral entre os genes) varia de 54,3% a 64,6%, com valor médio de 60,6%. O conteúdo GC (média geral entre os genes) varia de 35,4% a 45,7%, com um valor médio de 39,4% (Tabela 4). Em *Homo sapiens*, o conteúdo GC é 52,27%.

Tabela 2 – Valores RSCU dos genes E1, E2, E5, E6, E7, L1 e L2 (valores em negrito representam o códon preferencial)

Aminoácido	Códon	E1	E2	E5	E6	E7	L1	L2	Média
Ala	GCA	2,26	1,82	1,77	2,00	1,79	1,44	1,56	1,81
	GCC	0,42	0,79	0,35	0,35	0,70	0,83	0,67	0,59
	GCG	0,32	0,35	0,36	0,41	0,36	0,14	0,24	0,31
	GCU	1,01	1,04	1,28	1,23	1,16	1,59	1,53	1,26
Cys	UGC	0,51	0,73	0,65	0,80	0,64	0,51	0,91	0,68
	UGU	1,50	1,26	1,20	1,20	1,36	1,49	1,08	1,30
Asp	GAC	0,69	1,08	0,35	0,83	1,09	0,60	0,59	0,75
	GAU	1,31	0,92	0,75	1,17	0,91	1,40	1,41	1,13
Glu	GAA	1,32	1,38	0,37	1,34	0,93	1,36	1,37	1,15
	GAG	0,68	0,62	0,54	0,66	1,08	0,64	0,63	0,69
Phe	UUC	0,20	0,22	0,15	0,33	0,34	0,21	0,16	0,23
	UUU	1,80	1,78	1,82	1,67	1,21	1,79	1,84	1,70
Gly	GGA	1,48	1,56	0,71	1,51	1,68	0,87	0,79	1,23
	GGC	0,65	0,69	0,38	0,57	0,98	1,00	0,83	0,73
	GGG	0,92	0,78	0,67	1,03	0,56	0,52	0,74	0,74
	GGU	0,96	0,97	0,66	0,89	0,76	1,61	1,64	1,07
His	CAC	0,51	0,72	0,45	0,61	0,63	0,56	0,55	0,57
	CAU	1,49	1,28	1,51	1,37	1,23	1,44	1,45	1,40

Ile	AUA	1,72	1,45	1,48	1,51	1,03	1,09	1,23	1,36
	AUC	0,11	0,29	0,12	0,18	0,38	0,22	0,16	0,21
	AUU	1,17	1,26	1,31	1,31	1,59	1,69	1,60	1,42
Lys	AAA	1,57	1,53	0,40	1,29	0,94	1,26	1,45	1,21
	AAG	0,43	0,47	0,18	0,71	0,86	0,74	0,55	0,56
Leu	CUA	0,96	0,98	0,65	0,73	0,99	0,58	0,53	0,77
	CUC	0,07	0,31	0,04	0,09	0,21	0,12	0,11	0,14
	CUG	0,53	0,75	0,75	0,87	1,34	0,54	0,36	0,74
	CUU	0,46	0,84	0,92	0,67	1,00	0,77	0,89	0,79
	UUA	2,93	2,04	1,83	2,11	1,28	2,64	2,71	2,22
	UUG	1,05	1,08	1,81	1,52	1,19	1,35	1,39	1,34
Asn	AAC	0,45	0,91	0,29	0,65	0,74	0,54	0,42	0,57
	AAU	1,55	1,09	0,70	1,35	1,19	1,46	1,58	1,27
Pro	CCA	2,16	1,38	1,21	1,77	1,49	1,27	1,07	1,48
	CCC	0,26	0,93	0,40	0,41	0,64	0,58	0,60	0,55
	CCG	0,40	0,37	0,67	0,47	0,42	0,16	0,12	0,37
	CCU	1,18	1,32	1,40	1,34	1,46	1,98	2,21	1,56
Gln	CAA	1,21	1,26	0,90	1,13	1,21	1,05	1,26	1,14
	CAG	0,79	0,74	0,64	0,87	0,78	0,95	0,74	0,79
Arg	AGA	3,14	2,11	0,82	2,62	2,00	1,69	1,32	1,96
	AGG	1,00	0,89	0,66	1,12	0,71	1,32	0,98	0,95
	CGA	0,87	1,34	0,26	0,76	0,97	0,46	0,45	0,73
	CGC	0,21	0,30	0,37	0,49	0,61	0,94	1,04	0,56
	CGG	0,35	0,64	0,39	0,46	0,67	0,39	0,41	0,48
	CGU	0,43	0,71	1,57	0,56	1,04	1,19	1,81	1,04
Ser	AGC	1,01	0,78	0,25	0,66	1,36	0,47	0,51	0,72
	AGU	2,26	1,54	0,68	1,41	1,10	1,48	1,58	1,44
	UCA	1,27	1,15	0,73	1,30	1,72	0,97	0,96	1,16
	UCC	0,38	1,19	1,06	0,36	0,65	0,82	0,71	0,74
	UCG	0,22	0,32	0,41	0,54	0,31	0,16	0,20	0,31
	UCU	0,86	1,03	1,98	1,72	0,78	2,09	2,04	1,50
Thr	ACA	2,25	1,66	1,57	1,84	1,56	1,51	1,83	1,75
	ACC	0,34	1,07	1,08	0,53	0,94	0,75	0,64	0,77
	ACG	0,42	0,28	0,17	0,40	0,48	0,19	0,17	0,30
	ACU	0,99	0,98	0,70	1,22	1,02	1,55	1,36	1,12
Val	GUA	1,87	1,38	1,34	1,58	1,02	1,27	1,21	1,38
	GUC	0,14	0,29	0,28	0,21	0,41	0,27	0,31	0,27
	GUG	1,10	1,40	1,76	1,42	1,44	0,99	0,96	1,30
	GUU	0,89	0,93	0,62	0,80	1,14	1,48	1,53	1,05
Tyr	UAC	0,32	0,43	0,39	0,44	0,89	0,41	0,34	0,46
	UAU	1,68	1,57	1,58	1,56	0,83	1,59	1,66	1,50

Tabela 3 – Valores RSCU dos gêneros de HPV, representando a utilização de códons e frequência de utilização de códons em *Homo sapiens* (valores em negrito representam o códon preferencial)

Aminoácido	Códon	Alpha papillomavirus	Beta papillomavirus	Gamma papillomavirus	Mu papillomavirus	Nu papillomavirus	<i>Homo sapiens</i>
Ala	GCA	1,96	1,62	1,54	1,49	1,43	15,8
	GCC	0,66	0,57	0,46	0,53	0,67	27,7
	GCG	0,34	0,27	0,25	0,24	0,41	7,4

	GCU	1,02	1,53	1,75	1,74	1,49	18,4
Cys	UGC	0,66	0,88	0,56	0,77	0,65	12,6
	UGU	1,32	1,10	1,43	1,23	1,35	10,6
Asp	GAC	0,83	0,77	0,60	0,59	1,05	25,1
	GAU	1,04	1,23	1,40	1,41	0,95	21,8
Glu	GAA	1,14	1,27	1,34	1,22	0,87	29
	GAG	0,70	0,73	0,66	0,78	1,13	39,6
Phe	UUC	0,20	0,27	0,29	0,33	0,61	20,3
	UUU	1,68	1,71	1,70	1,45	1,39	17,6
Gly	GGA	1,20	1,32	1,43	1,24	0,90	16,5
	GGC	0,79	0,81	0,59	0,76	0,90	22,2
	GGG	0,76	0,72	0,69	0,49	1,19	16,5
	GGU	1,03	1,14	1,28	1,28	1,02	10,8
His	CAC	0,66	0,47	0,43	0,61	0,87	15,1
	CAU	1,32	1,49	1,52	1,39	1,13	10,9
Ile	AUA	1,47	1,16	1,16	1,07	1,40	7,5
	AUC	0,17	0,33	0,24	0,52	0,40	20,8
	AUU	1,34	1,51	1,60	1,41	1,20	16
Lys	AAA	1,18	1,37	1,41	1,16	1,16	24,4
	AAG	0,59	0,63	0,54	0,62	0,84	31,9
Leu	CUA	0,88	0,65	0,62	0,66	1,03	7,2
	CUC	0,10	0,27	0,15	0,30	0,54	19,6
	CUG	0,78	0,75	0,55	0,79	1,74	39,6
	CUU	0,66	0,98	0,98	1,13	0,86	13,2
	UUA	2,24	2,16	2,39	2,00	0,82	7,7
	UUG	1,34	1,19	1,31	1,12	1,01	12,9
Asn	AAC	0,63	0,57	0,47	0,49	0,66	19,1
	AAU	1,22	1,38	1,53	1,51	1,01	17
Pro	CCA	1,56	1,42	1,37	1,37	1,42	16,9
	CCC	0,62	0,54	0,38	0,50	0,78	19,8
	CCG	0,39	0,30	0,29	0,19	0,45	6,9
	CCU	1,38	1,74	1,94	1,95	1,35	17,5
Gln	CAA	1,11	1,25	1,27	1,16	0,66	12,3
	CAG	0,83	0,75	0,73	0,84	1,34	34,2
Arg	AGA	1,61	2,28	3,05	2,86	1,38	12,2
	AGG	0,98	1,15	0,80	1,05	1,25	12
	CGA	0,70	0,86	0,84	0,92	0,53	6,2
	CGC	0,60	0,65	0,44	0,37	0,60	10,4
	CGG	0,61	0,35	0,23	0,26	1,12	11,4
	CGU	1,23	0,71	0,65	0,54	0,85	4,5
Ser	AGC	0,81	0,70	0,65	0,83	1,81	19,5
	AGU	1,50	1,47	1,46	1,58	1,06	12,1
	UCA	1,16	1,29	1,19	1,00	0,90	12,2
	UCC	0,74	0,74	0,60	0,62	0,97	17,7
	UCG	0,31	0,28	0,30	0,28	0,24	4,4
	UCU	1,35	1,47	1,80	1,68	1,31	15,2

Thr	ACA	1,89	1,68	1,44	1,61	1,39	15,1
	ACC	0,78	0,73	0,64	0,90	0,66	18,9
	ACG	0,35	0,24	0,25	0,21	0,46	6,1
	ACU	0,90	1,35	1,67	1,28	1,50	13,1
Val	GUA	1,49	1,23	1,25	0,99	0,97	7,1
	GUC	0,20	0,45	0,32	0,36	0,60	14,5
	GUG	1,46	1,04	0,86	1,18	1,50	28,1
	GUU	0,85	1,28	1,57	1,47	0,93	11
Tyr	UAC	0,45	0,58	0,43	0,34	0,89	15,3
	UAU	1,52	1,35	1,51	1,66	1,11	12,2

4.2 Propriedades de composição em HPV

Nos genes E1, E2, E6 e E7, A% é maior que T%, com média de 35,3%, 35,4%, 31,7% e 28,7% para A%, e 27,5%, 28,5%, 29,7% e 25,6% para T%. No gene L1, não há diferença entre A% e T%, com média de 29,9%. Já nos genes E5 e L2, T% é maior que A%, com médias de 43,3% e 29,8% para T%, e 21,3% e 27,5% para A%. Em quase todos os genes, exceto L2, G% é maior que C%, com médias de 22,2%, 18,6%, 20,2%, 22,7%, 25,2% e 21,0% para G%, e 15,0%, 17,6%, 15,2%, 16,0%, 20,5% e 19,2% para C%. No gene L2, C% é maior que G%, com média de 22,2% e 20,6%, respectivamente (Tabela 4).

Nos genes E1, E2, E6 e E7, A₃% é maior que todos os outros, com médias de 38,5%, 34,1%, 34,5% e 30,3%, enquanto que nos genes E5, L1 e L2, T₃% é maior, com médias de 37,0%, 41,3% e 43,8%, respectivamente. Os valores de GC₃% variam de 14,3% a 65,1%, com uma média geral entre todos os genes de 30,7%, muito abaixo do valor GC₃% do hospedeiro, que é 58,55%. Os valores de ENC variam de 27,14 a 61,00, com um valor médio de 45,67 (Tabela 4).

Tabela 4 – Médias gerais por gene da composição de nucleotídeos e número efetivo de códons

Gene	T%	C%	A%	G%	T ₃ %	C ₃ %	A ₃ %	G ₃ %	GC%	AT%	GC ₃ %	ENC (média)
E1	27,5	15,0	35,3	22,2	33,2	10,6	38,5	17,7	37,2	62,8	25,1	41,65
E2	28,5	17,6	35,4	18,6	23,6	20,1	34,1	22,2	36,2	63,9	33,5	49,86
E5	43,3	15,2	21,3	20,2	37,0	10,5	27,3	25,2	35,4	64,6	31,9	41,60
E6	29,7	16,0	31,7	22,7	32,3	13,2	34,5	20,0	38,6	61,4	31,4	46,33
E7	25,6	20,5	28,7	25,2	28,8	17,8	30,3	23,1	45,7	54,3	39,7	50,04
L1	29,9	19,2	29,9	21,0	41,3	14,3	28,0	16,4	40,2	59,8	28,2	45,86
L2	29,8	22,2	27,5	20,6	43,8	13,9	29,9	12,3	42,6	57,3	25,2	44,32
Média	30,6	18,0	30,0	21,5	34,3	14,3	31,8	19,6	39,4	60,6	30,7	45,67

4.3 Análise de correlação

Foi verificada uma alta correlação entre GC₁₂% e GC₃% nos genes E1, E2, E6, L1 e L2 ($r = 0,703/0,545/0,560/0,575/0,790$, $P < 0,01$) e uma correlação significativa em E5 e E7 ($r = 0,223$ e $0,220$, $P < 0,01$).

Os resultados da análise de correlação entre A%, T%, C%, G% e GC% com A₃%, T₃%, C₃%, G₃% e GC₃% mostram que, no geral, eles possuem correlação significativa entre si (Tabelas 5 a 11).

Comparando os valores de ENC com GC% e GC₃%, uma correlação altamente significativa é obtida em E1 (ENC e GC%=0,526 / ENC e GC₃%=0,698, $P < 0,01$), E2 (ENC e GC%=0,788 / ENC e GC₃%=0,807, $P < 0,01$), E6 (ENC e GC%=0,672 / ENC e GC₃%=0,706, $P < 0,01$), L1 (ENC e GC%=0,749 / ENC e GC₃%=0,833, $P < 0,01$) e L2 (ENC e GC%=0,741 / ENC e GC₃%=0,632, $P < 0,01$). Entretanto, a correlação não foi significativa no gene E5. Além disso, a correlação foi significativa entre ENC e GC₃% (0,218, $P < 0,01$) do gene E7, mas não foi significativa entre ENC e GC% no mesmo gene.

Tabela 5 – Análise de correlação entre os conteúdos gerais e de terceira posição do gene E1

	T ₃	C ₃	A ₃	G ₃	GC ₃
T	0,971**	-0,697**	-0,316**	-0,440**	-0,628**
C	-0,562**	0,791**	-0,403**	0,663**	0,853**
A	-0,133*	-0,242**	0,936**	-0,632**	-0,532**
G	-0,746**	0,537**	-0,011	0,657**	0,695**
GC	-0,764**	0,747**	-0,206**	0,782**	0,919**

*Correlação significativa no nível 0,05/ ** Correlação significativa no nível 0,01

Tabela 6 – Análise de correlação entre os conteúdos gerais e de terceira posição do gene E2

	T ₃	C ₃	A ₃	G ₃	GC ₃
T	0,916**	-0,670**	0,375**	-0,546**	-0,836**
C	-0,825**	0,758**	-0,490**	0,563**	0,901**
A	0,546**	-0,582**	0,635**	-0,710**	-0,801**
G	-0,744**	0,586**	-0,574**	0,784**	0,877**
GC	-0,767**	0,916**	-0,900**	0,903**	0,804**

** Correlação significativa no nível 0,01

Tabela 7 – Análise de correlação entre os conteúdos gerais e de terceira posição do gene E5

	T ₃	C ₃	A ₃	G ₃	GC ₃
T	0,460**	-0,326**	-0,100	0,032	-0,116
C	-0,241**	0,549**	0,314**	-0,479**	-0,204*
A	-0,166	0,060	0,611**	-0,531**	-0,544**
G	-0,120	-0,064	-0,747**	0,837**	0,822**

GC	-0,386**	0,354**	-0,545**	0,555**	0,773**
----	----------	---------	----------	---------	---------

*Correlação significativa no nível 0,05/ ** Correlação significativa no nível 0,01

Tabela 8 – Análise de correlação entre os conteúdos gerais e de terceira posição do gene E6

	T ₃	C ₃	A ₃	G ₃	GC ₃
T	0,954**	-0,594**	-0,613**	-0,513**	-0,685**
C	-0,793**	0,821**	0,230**	0,475**	0,810**
A	-0,489**	-0,088	0,902**	0,031	-0,003
G	-0,402**	0,247**	-0,035	0,730**	0,537**
GC	-0,833**	0,781**	0,221**	0,668**	0,903**

** Correlação significativa no nível 0,01

Tabela 9 – Análise de correlação entre os conteúdos gerais e de terceira posição do gene E7

	T ₃	C ₃	A ₃	G ₃	GC ₃
T	0,887**	-0,396**	-0,390**	-0,302**	-0,563**
C	-0,434**	0,632**	0,060	-0,035	0,400**
A	-0,416**	0,192**	0,813**	-0,368**	-0,147*
G	-0,249**	-0,248**	-0,350**	0,769**	0,491**
GC	-0,598**	0,286**	-0,280**	0,677**	0,796**

*Correlação significativa no nível 0,05/ ** Correlação significativa no nível 0,01

Tabela 10 – Análise de correlação entre os conteúdos gerais e de terceira posição do gene L1

	T ₃	C ₃	A ₃	G ₃	GC ₃
T	0,909**	-0,416**	-0,034	-0,286**	-0,379**
C	-0,150*	0,841**	-0,850**	0,726**	0,851**
A	-0,134*	-0,651**	0,965**	-0,815**	-0,791**
G	-0,319**	0,583**	-0,732**	0,908**	0,814**
GC	-0,213**	0,809**	-0,902**	0,904**	0,935**

*Correlação significativa no nível 0,05/ ** Correlação significativa no nível 0,01

Tabela 11 – Análise de correlação entre os conteúdos gerais e de terceira posição do gene L2

	T ₃	C ₃	A ₃	G ₃	GC ₃
T	0,888**	-0,172**	-0,379**	-0,148*	-0,145*
C	0,230**	0,772**	-0,877**	0,580**	0,665**
A	-0,239**	-0,748**	0,966**	-0,754**	-0,714**
G	-0,674**	0,428**	-0,182**	0,658**	0,468**
GC	-0,135*	0,850**	-0,785**	0,835**	0,792**

*Correlação significativa no nível 0,05/ ** Correlação significativa no nível 0,01

4.4 Análise de componentes principais (PCA)

Na análise de componentes principais, $f'1$ corresponde a 80,72% da variação total e $f'2$ corresponde a 7,36% no gene E1, no gene E2 $f'1$ corresponde a 57,57% da variação total e $f'2$ corresponde a 9,29%, no gene E5 $f'1$ corresponde a 27,65% da variação total e $f'2$ corresponde a 8,35%, no gene E6 $f'1$ corresponde a 41,35% da variação total e $f'2$ corresponde a 12,60%, no gene E7 $f'1$ corresponde a 21,87% da variação total e $f'2$ corresponde a 9,37%, no gene L1 $f'1$ corresponde a 70,93% da variação total e $f'2$ corresponde a 9,95% e no gene L2 $f'1$ corresponde a 71,96% da variação total e $f'2$ corresponde a 8,84%.

Para o gene E1, a gama dos valores de distribuição do gráfico de $f'1$ é de 0,504 a 0,973 e de $f'2$ é de -0,498 a 0,560, para o gene E2, a gama dos valores de distribuição de $f'1$ é de 0,320 a 0,917 e de $f'2$ é de -0,515 a 0,605, para o gene E5, a gama dos valores de distribuição de $f'1$ é de 0,031 a 0,796 e de $f'2$ é de -0,386 a 0,764, para o gene E6, a gama dos valores de distribuição de $f'1$ é de 0,156 a 0,899 e de $f'2$ é de -0,732 a 0,659, para o gene E7, a gama dos valores de distribuição de $f'1$ é de 0,016 a 0,742 e de $f'2$ é de -0,616 a 0,607, para o gene L1, a gama dos valores de distribuição de $f'1$ é de 0,096 a 0,942 e de $f'2$ é de -0,518 a 0,694 e para o gene L2, a gama dos valores de distribuição de $f'1$ é de 0,190 a 0,941 e de $f'2$ é de -0,506 a 0,662.

Nos gráficos, em relação ao padrão de uso de códons sinônimos, cada gênero forma um agrupamento, porém sobrepostos em algumas regiões, principalmente os gêneros *Beta*, *Gamma* e *Mupapillomavirus*. A única sequência de *Nupapillomavirus* segue um padrão diferente de todos os outros gêneros, na maioria dos genes (Fig. 5, 6 e 7).

Nos genes reguladores (E1 e E2), o uso de códons segue padrões de dispersão diferentes, porém concentrados (Fig. 5). No gene E2 (Fig. 5a), o padrão é um pouco mais disperso que no gene E1 (Fig. 5b).

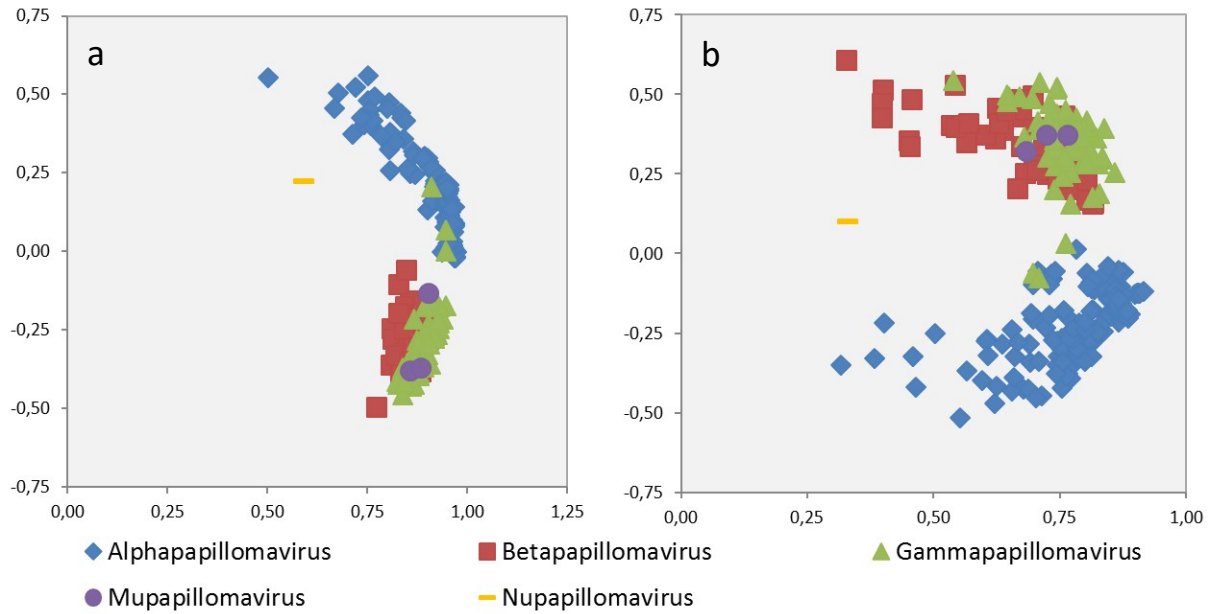


Fig. 5 – Padrão de uso de códons sinônimos (PCA) dos genes reguladores.

(a) Gene E1. (b) Gene E2

Nos oncogenes (E5, E6 e E7), o padrão é bastante disperso (Fig. 6), porém nos genes E6 e E7 nota-se um agrupamento discreto dos gêneros. Apesar das sequências do gene E5 serem exclusivamente do gênero *Alphapapillomavirus*, este apresenta um padrão totalmente disperso no gráfico (Fig. 6a).

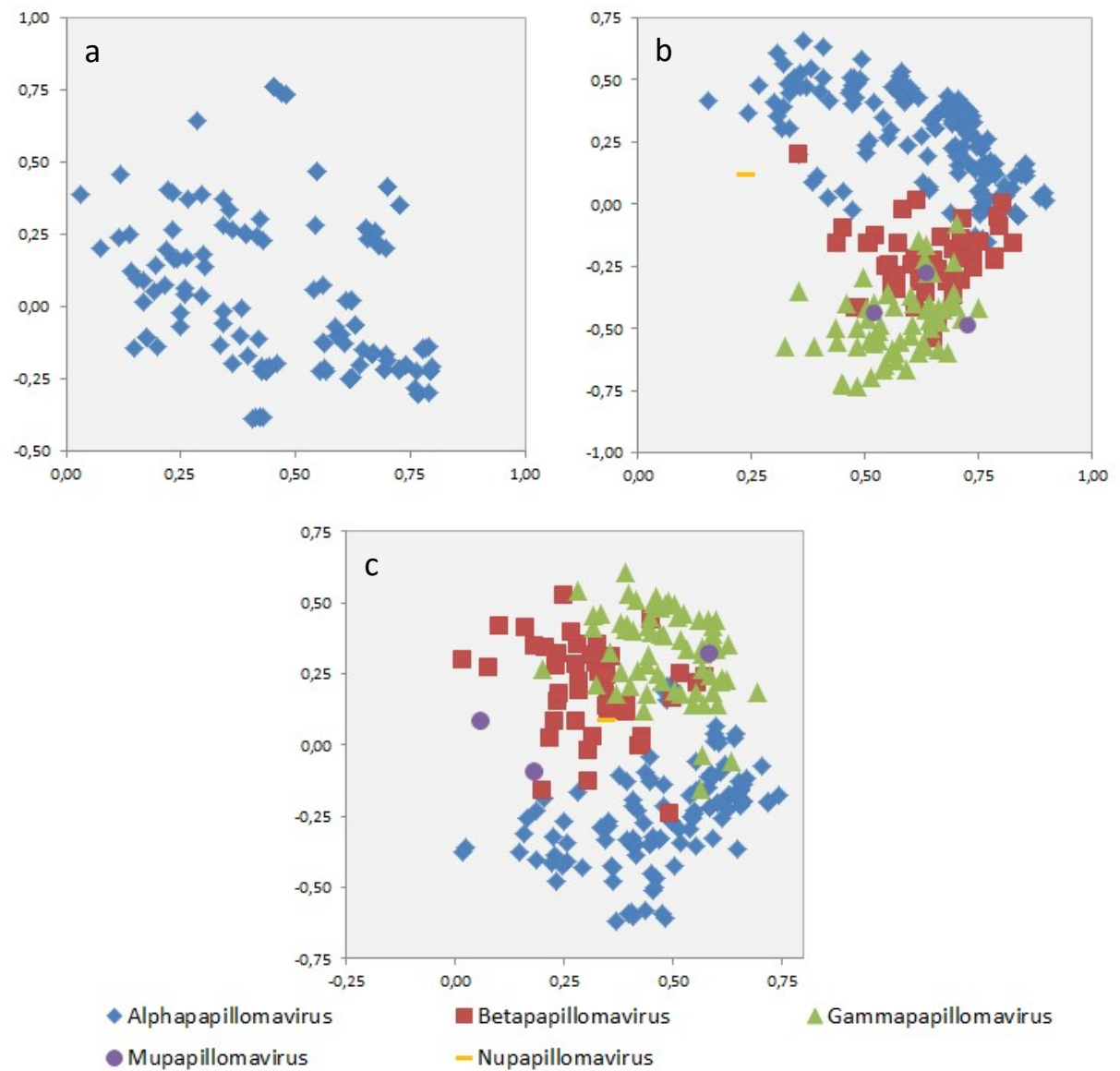


Fig. 6 – Padrão de uso de códons sinônimos (PCA) dos oncogenes.

(a) Gene E5. (b) Gene E6. (c) Gene E7.

Nos genes estruturais (L1 e L2), o padrão de dispersão é bastante semelhante e os agrupamentos são mais concentrados (Fig. 7).

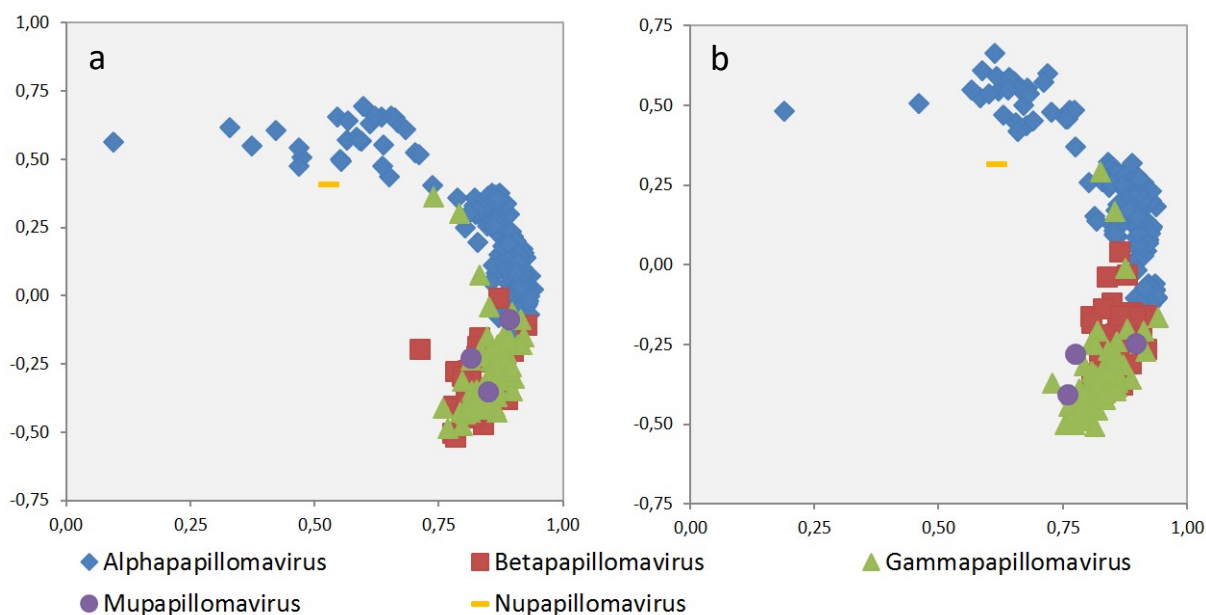


Fig. 7 – Padrão de uso de códons sinônimos (PCA) dos genes estruturais.
(a) Gene L1. (b) Gene L2.

Análise de correlação entre os eixos (índices de utilização de códon - $f'1$ e $f'2$) de PCA e aromaticidade e hidrofobicidade (GRAVY) mostrou uma significativa correlação negativa na maioria dos genes, conforme mostra a Tabela 12, e uma correlação significativa foi encontrada na análise entre os índices $f'1$ e $f'2$ e T%, C%, A%, G%, T₃%, C₃%, A₃%, G₃%, GC% e GC₃% (Tabela 13).

Tabela 12 – Análise de correlação de $f'1$ e $f'2$ e hidrofobicidade e aromaticidade

Gene	Índice	GRAVY	Aromaticidade
E1	$f'1$	0,072	-0,356**
	$f'2$	-0,241**	-0,642**
E2	$f'1$	-0,388**	-0,024
	$f'2$	0,125*	0,020
E5	$f'1$	-0,314**	-0,274**
	$f'2$	0,081	0,045
E6	$f'1$	0,396**	0,173**
	$f'2$	-0,054	-0,102
E7	$f'1$	0,122	-0,133*
	$f'2$	-0,072	-0,009
L1	$f'1$	-0,383**	-0,178**
	$f'2$	-0,018	0,220**
L2	$f'1$	-0,134*	-0,282**
	$f'2$	-0,157**	-0,147*

*Correlação significativa no nível 0,05

**Correlação significativa no nível 0,01

Tabela 13 – Análise de correlação de $f^{\circ}1$ e $f^{\circ}2$ e conteúdos

Gene	Índice	T%	C%	A%	G%	T ₃ %	C ₃ %	A ₃ %	G ₃ %	GC%	GC ₃ %
E1	$f^{\circ}1$	-0,081	-0,551**	0,835**	-0,058	-0,077	-0,334**	0,823**	-0,457**	-0,300**	-0,461**
	$f^{\circ}2$	-0,946**	0,489**	0,183**	0,709**	-0,953**	0,669**	0,352**	0,381**	0,690**	0,563**
E2	$f^{\circ}1$	-0,008	0,054	-0,169**	0,133*	-0,058	0,347**	-0,521**	0,347**	0,391**	0,091
	$f^{\circ}2$	0,029	-0,029	0,086	-0,103	0,050	-0,056	0,135*	-0,163**	-0,130*	-0,077
E5	$f^{\circ}1$	-0,355**	-0,075	0,115	0,278**	-0,208*	0,185*	-0,134	0,157	0,228**	0,215*
	$f^{\circ}2$	0,115	-0,007	0,171*	-0,260**	0,100	-0,100	0,167	-0,158	-0,249**	-0,224**
E6	$f^{\circ}1$	0,141*	-0,003	-0,383**	0,225**	0,117	0,175**	-0,392**	0,063	0,031	0,108
	$f^{\circ}2$	-0,165**	0,169**	0,050	0,162**	-0,164**	0,126*	0,033	0,230**	0,207**	0,207**
E7	$f^{\circ}1$	-0,231**	0,048	-0,160*	0,427**	-0,273**	-0,031	-0,203**	0,490**	0,445**	0,406**
	$f^{\circ}2$	-0,095	0,125	0,056	-0,042	-0,106	-0,012	0,054	0,047	0,048	0,060
L1	$f^{\circ}1$	-0,395**	-0,023	0,105	0,128*	-0,331**	0,081	0,076	0,023	0,041	0,052
	$f^{\circ}2$	0,058	-0,242**	0,121*	-0,024	0,029	-0,153*	0,127*	-0,072	-0,155**	-0,131*
L2	$f^{\circ}1$	-0,494**	-0,075	0,011	0,477**	-0,468**	0,186**	0,064	0,225**	0,217**	0,276**
	$f^{\circ}2$	-0,023	0,044	0,047	-0,098	-0,075	0,076	0,045	-0,100	-0,048	0,010

*Correlação significativa no nível 0,05/ ** Correlação significativa no nível 0,01

4.5 Análise de agrupamentos

Como resultado da análise de agrupamentos foram gerados dendrogramas para o padrão de uso de códons sinônimos para cada um dos genes dos HPVs. No gene E1, o dendrograma está dividido em três grandes agrupamentos (Fig. 8): o primeiro é formado por tipos pertencentes ao gênero *Alpha* e dois tipos do gênero *Gamma* (Fig. 8a), o segundo por todos os outros tipos de *Gamma* intercalados por *Beta* e todos os tipos de *Mu* (Fig. 8b) e o terceiro por *Alpha*, *Nu* e um tipo de *Betapapillomavirus* (Fig. 8c).

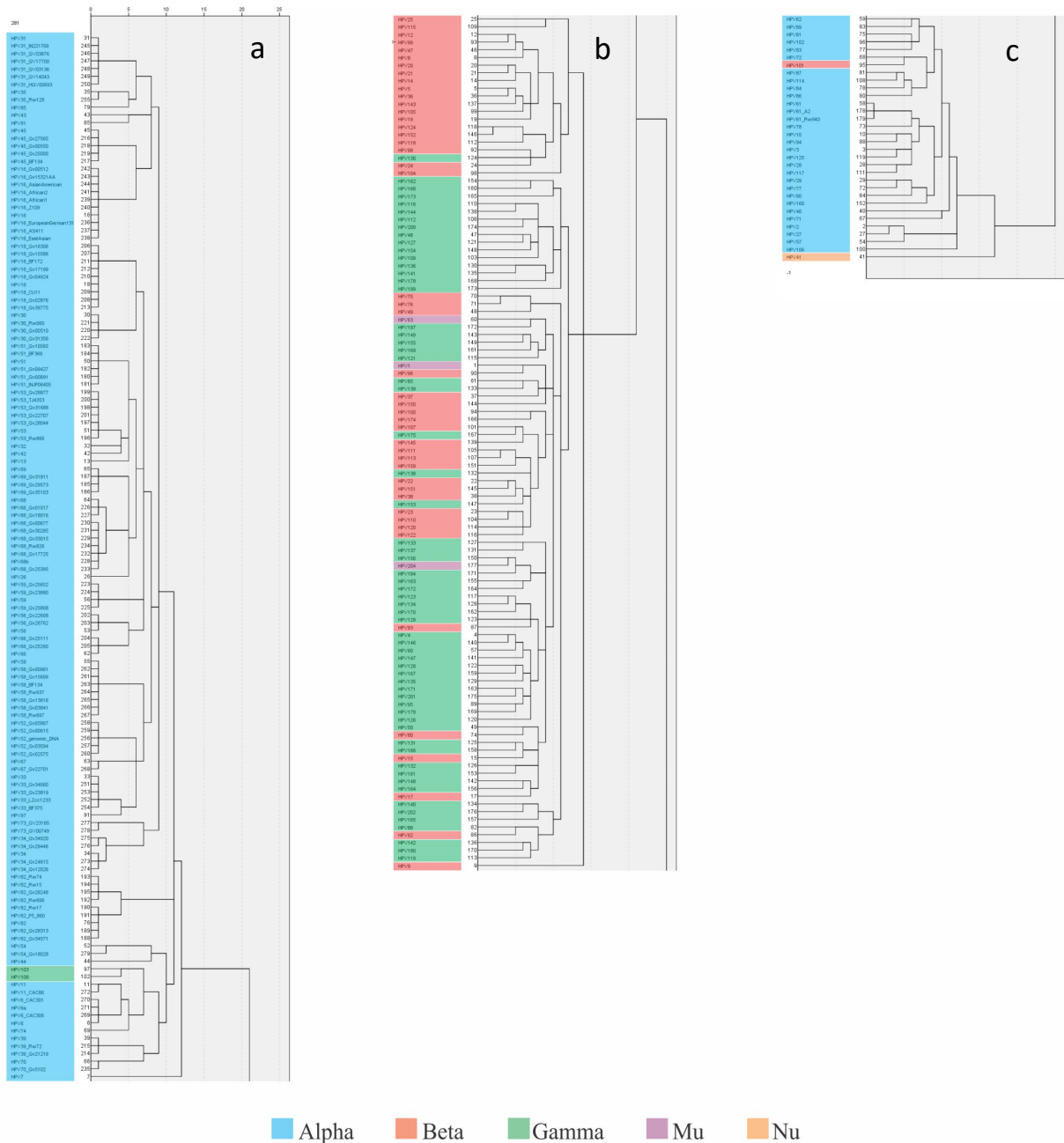


Fig. 8 – Análise de agrupamentos das relações de evolução na extensão da variação do uso de códons sinônimos do gene E1. (a) Primeiro agrupamento. (b) Segundo agrupamento. (c) Terceiro agrupamento.

No gene E2, há dois grandes agrupamentos formados (Fig. 9): o primeiro por todos os tipos do gênero *Alpha* e três tipos do gênero *Gamma* (Fig. 9a) e o segundo por todos os demais tipos do gênero *Gamma* e todos os tipos dos gêneros *Beta* e *Mu* (Fig. 9b). *Nupapillomavirus* está localizado separadamente de todos os outros.

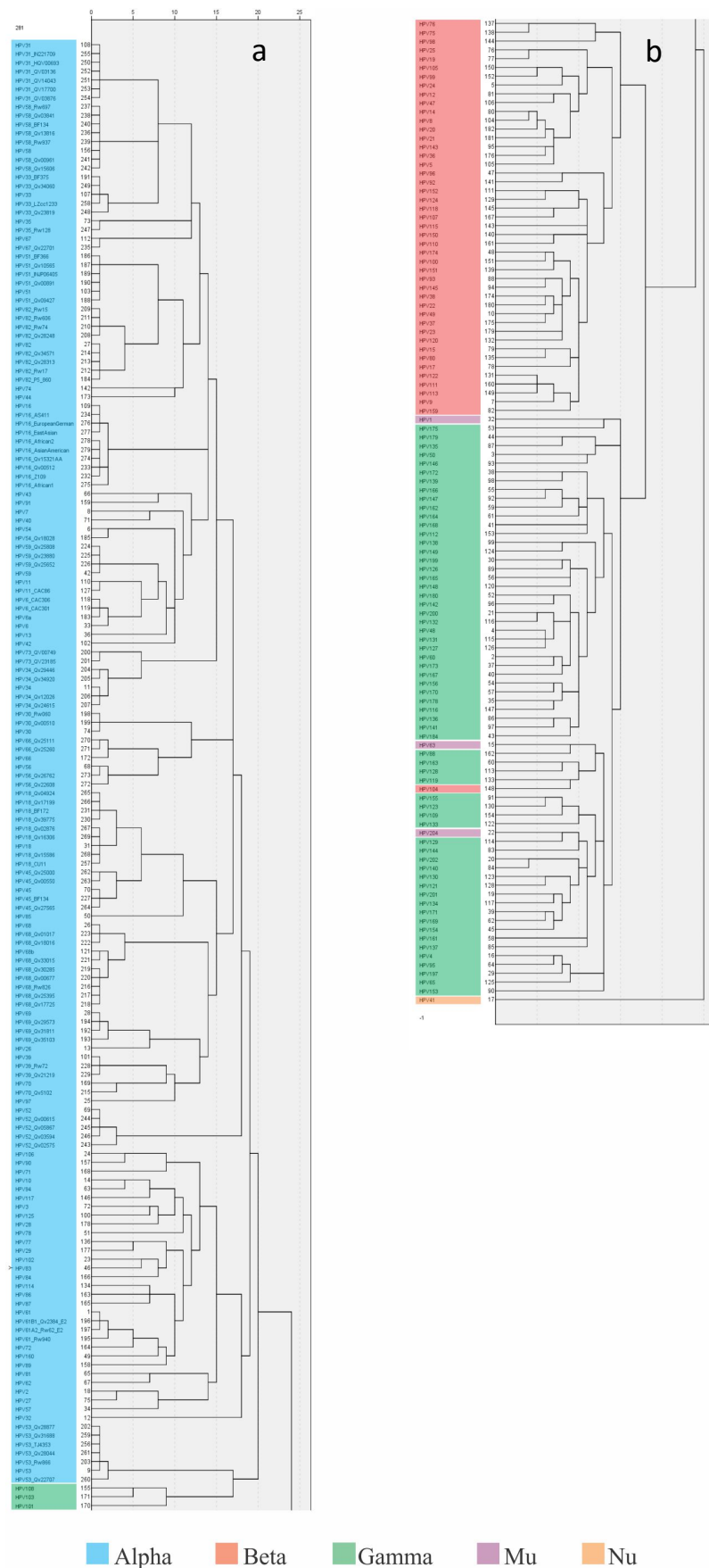


Fig. 9 – Análise de agrupamentos das relações de evolução na extensão da variação do uso de códons sinônimos do gene E2. (a) Primeiro agrupamento. (b) Segundo agrupamento.

O gene E5 não possui agrupamentos bem definidos (Fig. 10).

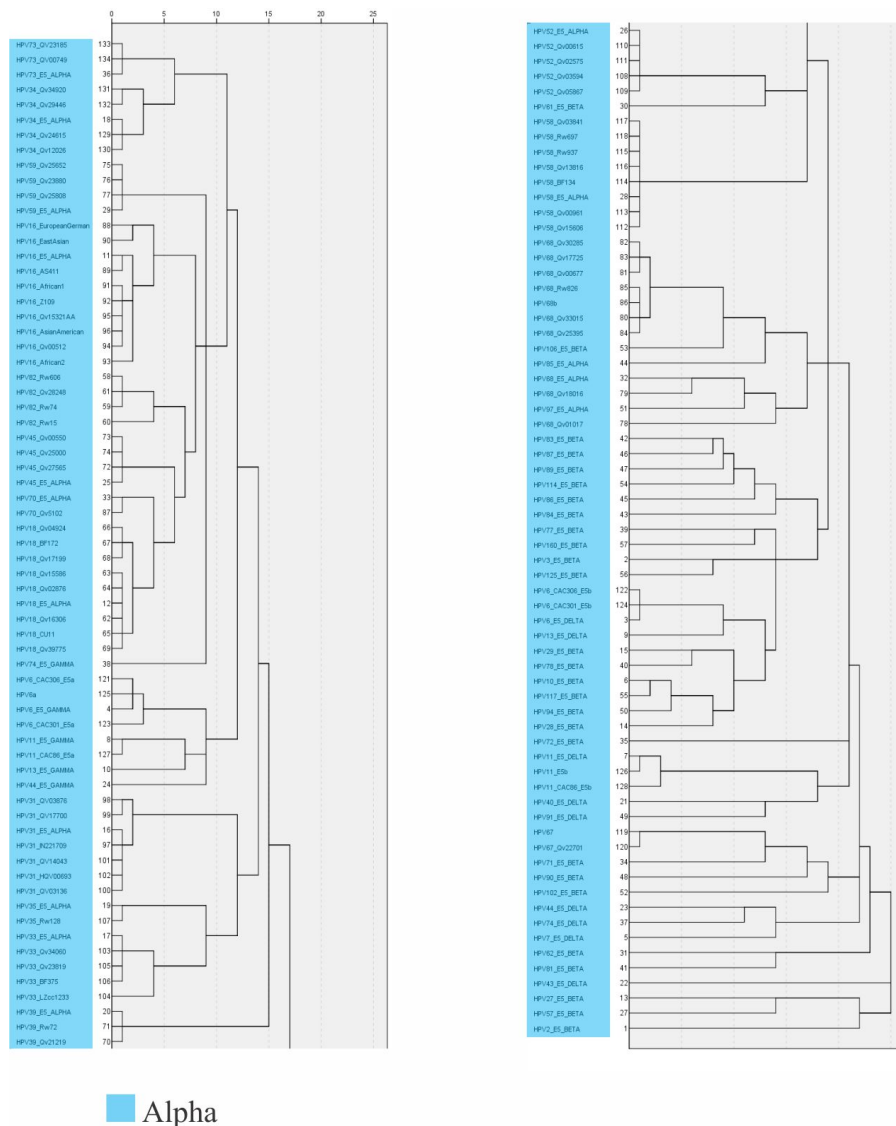


Fig. 10 – Análise de agrupamentos das relações de evolução na extensão da variação do uso de códons sinônimos do gene E5.

O gene E6 possui dois grandes agrupamentos (Fig. 11): o primeiro formado por *Alpha* intercalado com *Beta*, um tipo de *Gamma* e *Nu* (Fig. 11a), o segundo formado por *Gamma*, *Beta* e *Mu* intercalados (Fig. 11b). Três tipos de *Beta* e dois de *Gamma* estão separados dos agrupamentos maiores (Fig. 11c).

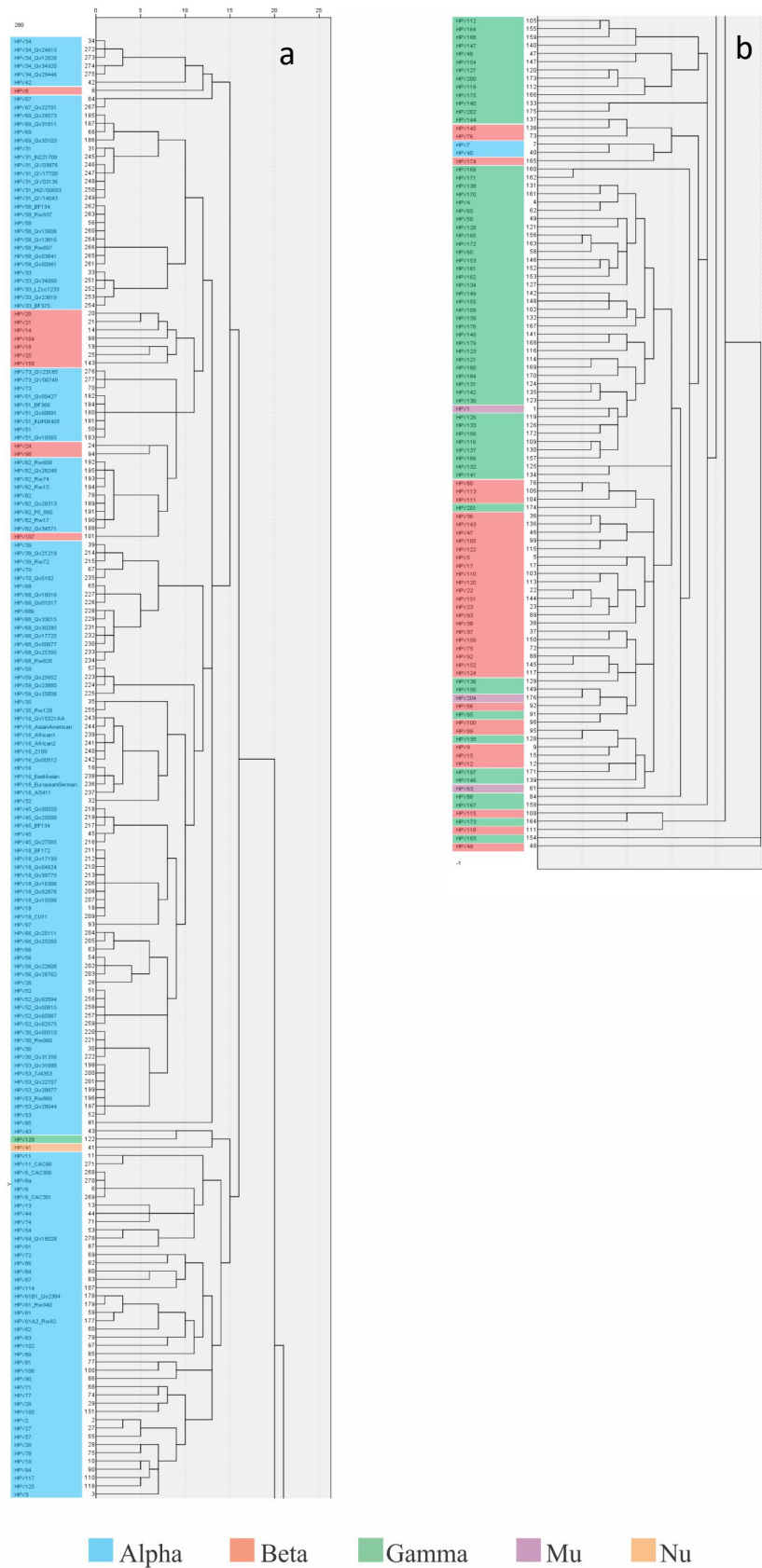


Fig. 11 – Análise de agrupamentos das relações de evolução na extensão da variação do uso de códons sinônimos do gene E6. (a) Primeiro agrupamento. (b) Segundo agrupamento.

No gene E7, o dendrograma não está bem agrupado. Há três grupos de *Alpha* e um grande grupo de *Gamma*, os demais gêneros estão intercalados (Fig. 12).



Fig. 12 – Análise de agrupamentos das relações de evolução na extensão da variação do uso de códons sinônimos do gene E7.

O gene L1 está dividido em três agrupamentos (Fig. 13): o primeiro formado por *Alpha*, dez tipos de *Gamma* e dois tipos de *Beta* (Fig. 13a), o segundo por *Gamma*, *Beta* e *Mu* intercalados (Fig. 13b) e o terceiro por *Alpha* e *Nu* (Fig. 13c).

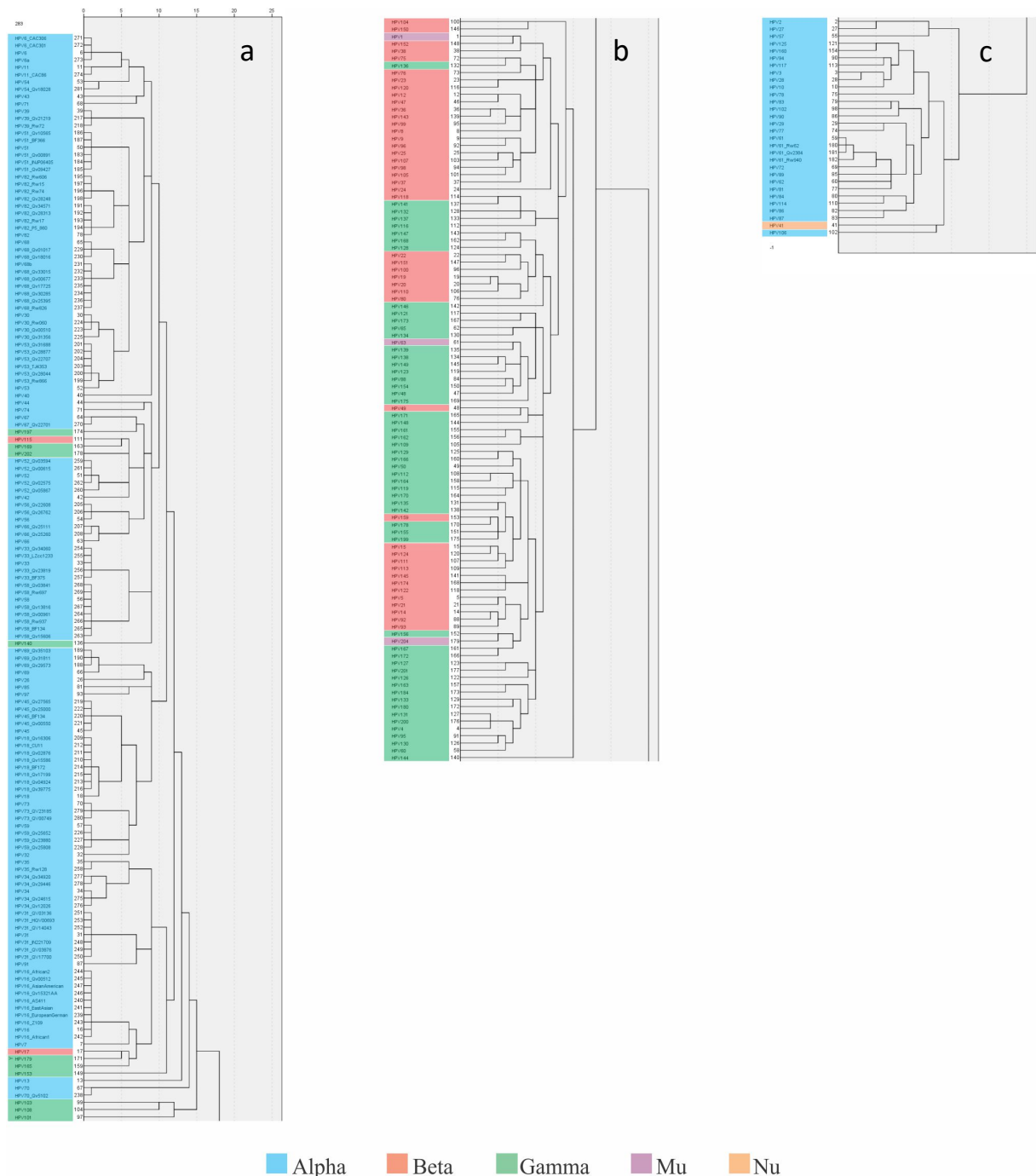


Fig. 13 – Análise de agrupamentos das relações de evolução na extensão da variação do uso de códons sinônimos do gene L1. (a) Primeiro agrupamento. (b) Segundo agrupamento. (c) Terceiro agrupamento.

O dendrograma do gene L2 está agrupado em três grupos (Fig. 14): o primeiro formado por *Alpha*, seis tipos de *Gamma* e um tipo de *Beta* (Fig. 14a), o segundo por *Gamma*, *Beta* e *Mu* intercalados (Fig. 14b) e o terceiro por *Alpha* e *Nu* (Fig. 14c).

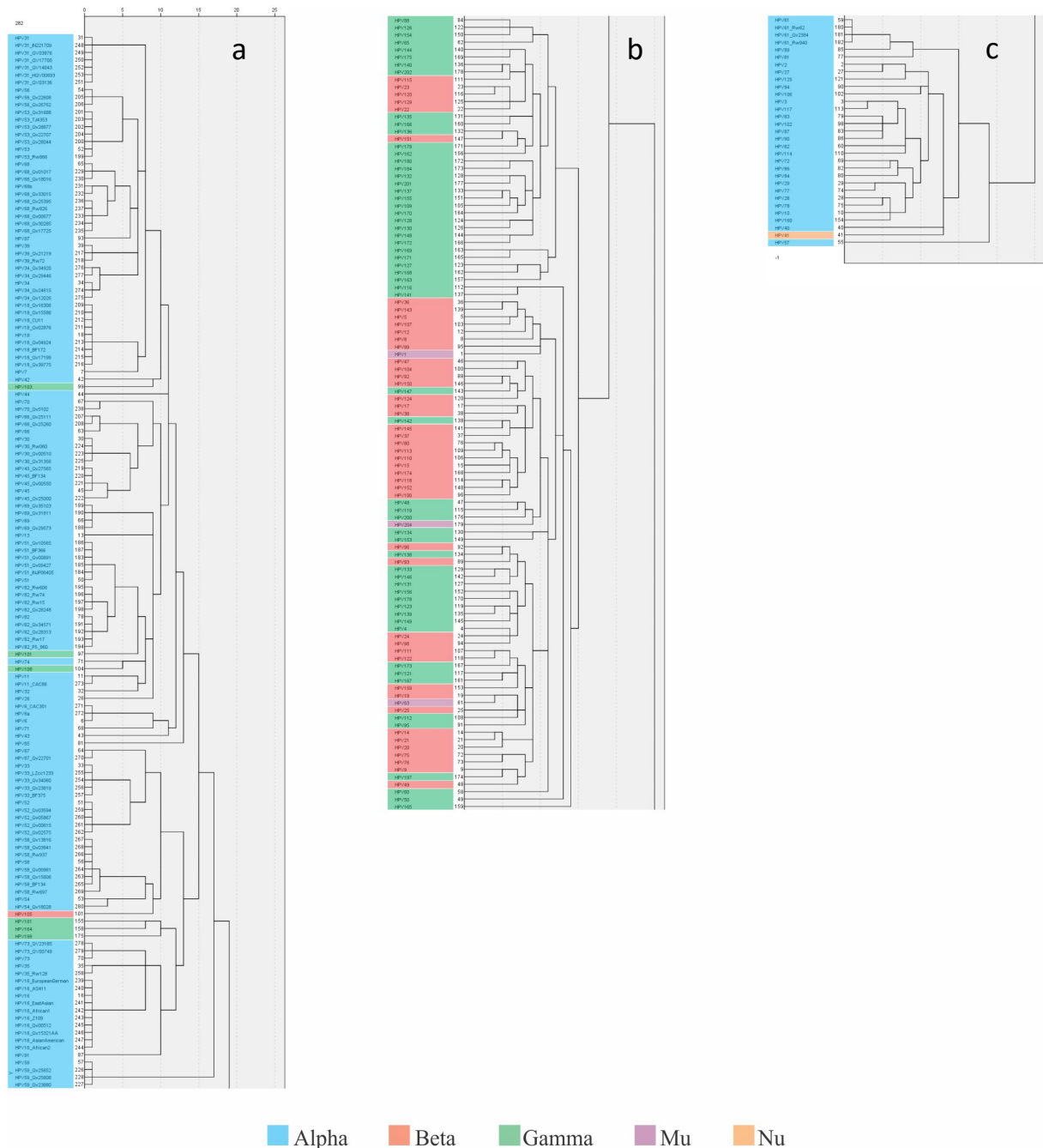


Fig. 14 – Análise de agrupamentos das relações de evolução na extensão da variação do uso de códons sinônimos do gene L2. (a) Primeiro agrupamento. (b) Segundo agrupamento. (c) Terceiro agrupamento.

5. DISCUSSÃO

Um alto conteúdo AT foi encontrado em HPV, diferentemente do genoma humano, que possui um alto conteúdo GC, assim como no estudo de Zhao, Liu & Frazer, (2003), o que leva a preferências de uso de códons sinônimos não congruentes entre o vírus e o hospedeiro, como observado também por Gottschiling et al. (2007). Isso pode ser explicado através de uma atualização da teoria da coevolução, que estabelece que a evolução do parasita

segue a evolução dos recursos do hospedeiro, não a evolução do próprio hospedeiro (VAN DOORSLAER, 2013). Além disso, existe a possibilidade da ocorrência da seleção de uso de códons para diminuir a síntese de proteína viral, reduzindo a exposição imune (VAN DOORSLAER, 2013; BRAVO & FÉLEZ-SÁNCHEZ, 2015; FÉLEZ-SÁNCHEZ et al., 2015). Uma outra hipótese seria que as preferências de utilização de códons dos HPVs podem ter evoluído para coincidir com o perfil variável do tRNA nos diferentes estágios de desenvolvimento de queratinócitos (células onde ocorre a replicação viral) (ZHAO & CHEN, 2011; BRAVO & FÉLEZ-SÁNCHEZ, 2015; FÉLEZ-SÁNCHEZ et al., 2015), sendo influenciado também pelo tropismo do vírus, ou seja, sua apresentação clínica (CLADEL, BERTOTTO & CHRISTENSEN, 2010).

Os conteúdos $A_3\%$ e $T_3\%$ têm uma frequência muito maior que $C_3\%$ e $G_3\%$, o que sugere que o conteúdo de nucleotídeos na terceira posição dos códons não é determinado por toda a composição de nucleotídeos (XU et al., 2015), o mesmo resultado foi encontrado por Zhao, Liu & Frazer (2003), que afirmam que há a possibilidade desse alto conteúdo estar relacionado com a adaptabilidade genética específica de HPV. Os valores de ENC mostram que existe um significativo viés no uso de códons sinônimos em HPV. É interessante notar que o valor ENC é menor que o de outros vírus de DNA, que variam de 48,30 a 56,80 (SHACKELTON, PARRISH & HOLMES, 2006). Portanto, o viés de uso de códons em HPV é maior que em outros vírus. Esse uso desigual de códons indica que substituições sinônimas de códons em HPV não são neutras nem aleatórias (ZHAO, LIU & FRAZER, 2003), mas, provavelmente, causadas por fatores externos, como a pressão de mutação e a seleção de tradução.

Apesar da taxa de mutação do vírus DNA não ser maior do que os vírus de RNA, a pressão de mutação é o principal fator de variação de utilização de códons (SHACKELTON, PARRISH & HOLMES, 2006). Pressão de mutação é a ocorrência repetida de erros de mutação específicos durante a replicação do genoma, que é afetada pelas restrições de composição genética (DURET, 2002). A correlação altamente significativa entre $GC_{12}\%$ e $GC_3\%$ mostra que o principal fator influenciador é a pressão de mutação, que ocorre também em outros vírus de DNA (XU et al., 2015). Além disso, a correlação significativa entre $A\%$, $T\%$, $C\%$, $G\%$, $GC\%$ e $A_3\%$, $T_3\%$, $C_3\%$, $G_3\%$, $GC_3\%$ junto com a correlação significativa entre ENC e $GC\%$ e entre ENC e $GC_3\%$ indicam que as restrições de composição influenciam a variação do padrão de uso de códons em HPV, corroborando com o resultado anterior.

Se a seleção de tradução contribui para o viés do padrão de uso de códons, os valores ENC dos genes estruturais serão maiores que os dos genes não estruturais (ZHOU et al., 2005). Neste estudo, o valor médio de ENC dos genes estruturais é de 45,09 e dos genes não estruturais 45,90, ou seja, não há uma diferença significativa, o que significa que a seleção de tradução não é um fator determinante para a preferência de uso de códons dos HPVs.

A correlação significativa entre f^1 , f^2 e aromaticidade e hidrofobicidade mostra que a aromaticidade e a hidrofobicidade dos aminoácidos também influenciam na variação do padrão de uso de códons. Códons com um alto conteúdo AT, geralmente, codificam aminoácidos hidrofóbicos (D'ONOFRIO et al., 1999) e aromaticidade pode constituir, até certo ponto, maior estabilidade nas ligações entre A e T (GUILLAUMES, SIMON & GUERRA, 2015).

Os resultados obtidos em PCA mostram que os diferentes gêneros (especialmente *Alphapapillomavirus*) se apresentam agrupados, na maioria dos genes, o que sugere que os gêneros estão relacionados com a variação de utilização de códons sinônimos, apesar de os agrupamentos dos gêneros *Gamma*, *Beta* e *Mu* estarem parcialmente sobrepostos, mostrando padrões de uso de códons muito semelhantes entre eles. Cladel, Bertotto & Christensen (2010) e Féléz-Sánchez et al. (2015) também encontraram diferenças no uso de códons entre o gênero *Alpha* e os demais, e concluíram que a utilização de códons nos papilomavírus pode desempenhar um papel na facilitação da fuga da vigilância imunológica, assim como o tropismo e a produtividade podem ser responsáveis pela variação na preferência de uso de códons.

Conforme a variação dos valores de distribuição, percebe-se que os genes não estruturais são mais conservados que os genes estruturais, pois exibem menor variação nesses valores. Isso pode ser explicado pelo fato de que as proteínas estruturais são aquelas que interagem diretamente com os receptores das células do hospedeiro para infectar essas células, e consequentemente, elas interagem mais com o sistema imune do hospedeiro. Desta forma, a maior variação no uso de códons pode ser devido a este mecanismo de escape imunológico. Eventos de recombinação podem ter ocorrido ao longo da história evolutiva causando as variações observadas no uso de códons entre os gêneros e os genes (SHAH, DOORBAR & GOLDSTEIN, 2010).

Assim como nos gráficos de dispersão, os dendrogramas mostram uma separação clara entre o gênero *Alphapapillomavirus* e os demais, mostrando que este possui um padrão diferente de uso de códons sinônimos. Os outros gêneros têm padrões semelhantes entre si, o

que corrobora com o resultado obtido em PCA. O gênero *Alpha* possui os tipos de HPV de alto risco, e essa utilização diferenciada de códons pode estar relacionada com a sua adaptação ao hospedeiro permitindo aos vírus deste gênero graus maiores de patogenicidade.

6. CONCLUSÃO

Este estudo revelou que o viés de uso de códons é alto em HPV, e a pressão de mutação é o principal fator que afeta a variação do uso de códons. Outros fatores, incluindo restrições de composição, aromaticidade e hidrofobicidade também podem, em menor escala, controlar conjuntamente o viés de uso de códons sinônimos. Além disso, a análise dos padrões com base no método de PCA indica que gêneros apresentam perfis de uso de códons, especialmente o gênero *Alpha*, que possui os tipos de HPV de alto risco, indicando maior adaptação ao hospedeiro. Foi verificado que o padrão de utilização de códons de HPV é diferente do padrão de utilização de códons de seu hospedeiro. Com a utilização das sequências das variantes de HPV, neste estudo, os resultados de outros autores em trabalhos anteriores foram confirmados. Entretanto, outros estudos ainda precisam ser realizados para que se determine as possíveis relações entre os padrões de uso de códons em HPV, sua diversificação, e a associação com os diferentes níveis de patogenicidade dos diferentes tipos e variantes virais.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ALBERTS, B. et al. **Fundamentos da Biologia Celular: uma introdução à biologia molecular da célula**. 1 ed., Porto Alegre: Artmed, 1999.

ARBYN, M. et al. Are 20 human papillomavirus types causing cervical cancer?. **The Journal of pathology**, v. 234, n. 4, p. 431-435, 2014.

BERNARD, H.-U. et al. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. **Virology**, v. 401, n. 1, p. 70-79. Elsevier, 2010.

BERNARD, H.-U.; CALLEJA-MACIAS, I. E.; DUNN, S. T. Genome variation of human papillomavirus types: phylogenetic and medical implications. **International journal of cancer**, v. 118, n. 5, p. 1071-1076, 2006.

BICHO, M. C. Biomarkers of cervical carcinogenesis associated with genital HPV infection. **Acta Médica Portuguesa**, v. 26, n. 2, p. 79-80, 2013.

BRAVO, I. G.; DE SANJOSÉ, S.; GOTTSCHLING, M. The clinical importance of understanding the evolution of papillomaviruses. **Trends in Microbiology**, v. 18, n. 10, p. 432-438, 2010.

BRAVO, I. G.; FÉLEZ-SÁNCHEZ, M. Papillomaviruses: viral evolution, cancer and evolutionary medicine. **Evolution, Medicine, and Public Health**, v. 2015, n. 1, p. 32-51, 2015.

BRAVO, I. G.; MÜLLER, M. Codon usage in papillomavirus genes: practical and functional aspects. **Papillomavirus Report**, v. 16, n. 2, p. 63-72, 2005.

CLADEL, N. M.; BERTOTTO, A.; CHRISTENSEN, N. D. Human alpha and beta papillomaviruses use different synonymous codon profiles. **Virus genes**, v. 40, n. 3, p. 329-340, 2010.

DE VILLIERS, E.-M. et al. Classification of papillomaviruses. **Virology**, v. 324, n. 1, p. 17-27. Elsevier, 2004.

D'ONOFRIO, G. et al. The correlation of protein hydropathy with the base composition of coding sequences. **Gene**, v. 238, n. 1, p. 3-14, 1999.

DOORBAR, J. et al. The biology and life-cycle of human papillomaviruses. **Vaccine**, v. 30, p. F55-F70. Elsevier, 2012.

DURET, L. Evolution of synonymous codon usage in metazoans. **Current opinion in genetics & development**, v. 12, n. 6, p. 640-649, 2002.

FÉLEZ-SÁNCHEZ, M. et al. Cancer, warts, or asymptomatic infections: clinical presentation matches codon usage preferences in human Papillomaviruses. **Genome Biology and Evolution**, v. 7, n. 8, p. 2117-2135, 2015.

GOTTSCHLING, M. et al. Multiple evolutionary mechanisms drive papillomavirus diversification. **Molecular Biology and Evolution**, v. 24, n. 5, p. 1242-1258, 2007.

GOTTSCHLING, M. et al. Quantifying the phylodynamic forces driving papillomavirus evolution. **Molecular Biology and Evolution**, v. 28, n. 7, p. 2101-2113, 2011.

GUILLAUMES, L.; SIMON, S.; GUERRA, C. F. The Role of Aromaticity, Hybridization, Electrostatics, and Covalency in Resonance-Assisted Hydrogen Bonds of Adenine–Thymine (AT) Base Pairs and Their Mimics. **ChemistryOpen**, v. 4, n. 3, p. 318-327, 2015.

INFORMATION CENTRE ON HPV AND CERVICAL CANCER. Brazil: human papillomavirus and related cancers – fact sheet 2014. **World Health Organization**, 2014.

INTERNATIONAL AGENCY FOR RESEARCH ON CANCER. **IARC monographs on the evaluation of carcinogenic to humans: human papillomavirus**, v. 90. Lyon: WHO Press, 2005.

MATTIOLI, M. A. P. et al. **Estudo do padrão de utilização de códons em sequencias genicas de eucalipto visando a maximização da produção de proteínas através da otimização das sequencias de DNA**. 2008. Tese (Mestrado em Genética e Biologia Molecular) – Universidade Estadual de Campinas, Campina, 2008.

SANGAR, V. C. et al. Human papillomavirus biology: review. **International Journal of Advanced research**, v. 2, n. 5, p. 125-133, 2014.

SANTOS, N. S. O.; ROMANOS, M. T. V.; WIGG, M. D. **Introdução à virologia humana**. 2. ed. Rio de Janeiro: Guanabara Koogan, 2008.

SHACKELTON, L. A.; PARRISH, C. R.; HOLMES, E. C. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. **Journal of molecular evolution**, v. 62, n. 5, p. 551-563, 2006.

SHAH, S. D.; DOORBAR, J.; GOLDSTEIN, R. A. Analysis of host–parasite incongruence in papillomavirus evolution using importance sampling. **Molecular Biology and Evolution**, v. 27, n. 6, p. 1301-1314, 2010.

STANLEY, M. A. Epithelial cell responses to infection with human papillomavirus. **Clinical microbiology reviews**, v. 25, n. 2, p. 215-222, 2012.

VAN DOORSLAER, K. Evolution of the papillomaviridae. **Virology**, v. 445, n. 1, p. 11-20, 2013.

WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. Chemometrics and intelligent laboratory systems, v. 2, n. 1, p. 37-52, 1987. In: XU, Y. et al. Analysis of synonymous codon usage pattern in duck circovirus. **Gene**, v. 557, n. 2, p. 138-145, 2015.

WORLD HEALTH ORGANIZATION. Comprehensive cervical cancer control: a guide to essential practice. 2. ed. **World Health Organization**, 2014.

WRIGHT, Frank. The 'effective number of codons' used in a gene. *Gene*, v. 87, n. 1, p. 23-29, 1990. In: XU, Y. et al. Analysis of synonymous codon usage pattern in duck circovirus. **Gene**, v. 557, n. 2, p. 138-145, 2015.

XU, Y. et al. Analysis of synonymous codon usage pattern in duck circovirus. **Gene**, v. 557, n. 2, p. 138-145, 2015.

ZHAO, K.-N.; CHEN, J. Codon usage roles in human papillomavirus. **Reviews in Medical Virology**, v. 21, n. 6, p. 397-411, 2011.

ZHAO, K.-N.; LIU, W. J.; FRAZER, I. H. Codon usage bias and A+ T content variation in human papillomavirus genomes. **Virus Research**, v. 98, n. 2, p. 95-104, 2003.

ZHOU, T. et al. Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. **Biosystems**, v. 81, n. 1, p. 77-86, 2005.

Apêndice A – Sequências utilizadas na pesquisa

Tipo	Variante	Gênero	Linhagem	Sublinhagem	Número de acesso <i>GenBank</i>
HPV1	-	Mu	-	-	V01116
HPV2	-	Alpha	-	-	NC_001352
HPV3	-	Alpha	-	-	X74462
HPV4	-	Gamma	-	-	NC_001457
HPV5	-	Beta	-	-	M17463
HPV6	-	Alpha	A	A1	X00203
HPV6	CAC306	Alpha	B	B1	FR751337
HPV6	CAC301	Alpha	B	B2	FR751328
HPV6	B3	Alpha	B	B3	L41216
HPV7	-	Alpha	-	-	NC_001595
HPV8	-	Beta	-	-	M12737
HPV9	-	Beta	-	-	NC_001596
HPV10	-	Alpha	-	-	NC_001576
HPV11	-	Alpha	A	A1	M14119
HPV11	CAC86	Alpha	A	A2	FN907962
HPV12	-	Beta	-	-	X74466
HPV13	-	Alpha	-	-	X62843
HPV14	-	Beta	-	-	X74467
HPV15	-	Beta	-	-	X74468
HPV16	-	Alpha	A	A1	K02718
HPV16	European				
	German 131	Alpha	A	A2	AF536179
HPV16	AS411	Alpha	A	A3	HQ644236
HPV16	East Asian	Alpha	A	A4	AF534061
HPV16	African 1	Alpha	B	B1	AF536180
HPV16	Z109	Alpha	B	B2	HQ644298
HPV16	African 2	Alpha	C	C1	AF472509
HPV16	Qv00512	Alpha	D	D1	HQ644257
HPV16	Qv15321AA	Alpha	D	D2	AY686579
HPV16	Asian				
	American	Alpha	D	D3	AF402678
HPV17	-	Beta	-	-	X74469
HPV18	-	Alpha	A	A1	AY262282
HPV18	Qv16306	Alpha	A	A2	EF202146
HPV18	Qv15586	Alpha	A	A3	EF202147
HPV18	Qv02876	Alpha	A	A4	EF202151
HPV18	CU11	Alpha	A	A5	GQ180787
HPV18	Qv04924	Alpha	B	B1	EF202155
HPV18	BF172	Alpha	B	B2	KC470225
HPV18	Qv17199	Alpha	B	B3	EF202152
HPV18	Qv39775	Alpha	C	C1	KC470229
HPV19	-	Beta	-	-	X74470
HPV20	-	Beta	-	-	U31778
HPV21	-	Beta	-	-	U31779
HPV22	-	Beta	-	-	U31780

HPV23	-	Beta	-	-	U31781
HPV24	-	Beta	-	-	NC_001683.1
HPV25	-	Beta	-	-	X74471
HPV26	-	Alpha	-	-	X74472
HPV27	-	Alpha	-	-	X74473
HPV28	-	Alpha	-	-	U31783
HPV29	-	Alpha	-	-	U31784
HPV30	-	Alpha	A	A1	X74474
HPV30	Qv00510	Alpha	A	A2	KF436842
HPV30	Rw060	Alpha	A	A3	KF436844
HPV30	Qv31356	Alpha	B	B1	KF436850
HPV31	-	Alpha	A	A1	J04353
HPV31	IN221709	Alpha	A	A2	HQ537675
HPV31	QV03876	Alpha	B	B1	HQ537676
HPV31	QV17700	Alpha	B	B2	HQ537680
HPV31	QV03136	Alpha	C	C1	HQ537682
HPV31	QV14043	Alpha	C	C2	HQ537684
HPV31	HQV00693	Alpha	C	C3	HQ537685
HPV32	-	Alpha	-	-	NC_001586
HPV33	-	Alpha	A	A1	M12732
HPV33	Qv34060	Alpha	A	A2	HQ537698
HPV33	LZcc1233	Alpha	A	A3	EU918766
HPV33	Qv23819	Alpha	B	B1	HQ537705
HPV33	BF375	Alpha	C	C1	KF436865
HPV34	-	Alpha	A	A1	X74476
HPV34	Qv24615	Alpha	A	A2	KF436808
HPV34	Qv12026	Alpha	B	B1	KF436810
HPV34	Qv34920	Alpha	C	C1	KF436812
HPV34	Qv29446	Alpha	C	C2	KF436816
HPV35	-	Alpha	A	A1	X74477
HPV35	Rw128	Alpha	A	A2	HQ537727
HPV36	-	Beta	-	-	U31785
HPV37	-	Beta	-	-	U31786
HPV38	-	Beta	-	-	U31787
HPV39	-	Alpha	A	A1	M62849
HPV39	Qv21219	Alpha	A	A2	KC470239
HPV39	Rw72	Alpha	B	B1	KC470247
HPV40	-	Alpha	-	-	X74478
HPV41	-	Nu	-	-	NC_001354
HPV42	-	Alpha	-	-	M73236
HPV43	-	Alpha	-	-	AJ620205
HPV44	-	Alpha	-	-	U31788
HPV45	-	Alpha	A	A1	X74479
HPV45	Qv27565	Alpha	A	A2	EF202157
HPV45	BF134	Alpha	A	A3	KC470256
HPV45	Qv00550	Alpha	B	B1	EF202161

HPV45	Qv25000	Alpha	B	B2	EF202164
HPV47	-	Beta	-	-	M32305
HPV48	-	Gamma	-	-	NC_001690
HPV49	-	Beta	-	-	NC_001591
HPV50	-	Gamma	-	-	NC_001691
HPV51	-	Alpha	A	A1	M62877
HPV51	Qv00891	Alpha	A	A2	KF436870
HPV51	INJP06405	Alpha	A	A3	KF436873
HPV51	Qv09427	Alpha	A	A4	KF436875
HPV51	Qv10565	Alpha	B	B1	KF436883
HPV51	BF366	Alpha	B	B2	KF436886
HPV52	-	Alpha	A	A1	X74481
HPV52	Qv03594	Alpha	B	B1	HQ537740
HPV52	Qv05867	Alpha	C	C1	HQ537744
HPV52	Qv00615	Alpha	C	C2	HQ537746
HPV52	Qv02575	Alpha	D	D1	HQ537748
HPV53	-	Alpha	A	A1	X74482
HPV53	Rw866	Alpha	B	B1	KF436818
HPV53	Qv28044	Alpha	C	C1	EF546477
HPV53	Qv31688	Alpha	D	D1	EF546482
HPV53	Qv28877	Alpha	D	D2	KF436823
HPV53	TJ4353	Alpha	D	D3	GQ472849
HPV53	Qv22707	Alpha	D	D4	EF546479
HPV54	-	Alpha	A	A1	U37488
HPV54	Qv18028	Alpha	C	C1	KF436894
HPV56	-	Alpha	A	A1	X74483
HPV56	Qv22608	Alpha	A	A2	EF177179
HPV56	Qv26762	Alpha	B	B1	EF177176
HPV57	-	Alpha	-	-	X55965
HPV58	-	Alpha	A	A1	D90400
HPV58	Qv15606	Alpha	A	A2	HQ537752
HPV58	Qv00961	Alpha	A	A3	HQ537758
HPV58	BF134	Alpha	B	B1	HQ537762
HPV58	Rw937	Alpha	B	B2	HQ537764
HPV58	Qv13816	Alpha	C	C1	HQ537774
HPV58	Qv03841	Alpha	D	D1	HQ537768
HPV58	Rw697	Alpha	D	D2	HQ537770
HPV59	-	Alpha	A	A1	X77858
HPV59	Qv25652	Alpha	A	A2	KC470261
HPV59	Qv23880	Alpha	A	A3	KC470263
HPV59	Qv25808	Alpha	B	B1	KC470264
HPV60	-	Gamma	-	-	NC_001693
HPV61	-	Alpha	A	A1	U31793
HPV61	Rw940	Alpha	A	A2	KF436853
HPV61	Qv2384	Alpha	B	B1	KF436856
HPV61	Rw62	Alpha	C	C1	KF436858

HPV62	-	Alpha	-	-	AY39570
HPV63	-	Mu	-	-	NC_001458
HPV65	-	Gamma	-	-	X70829
HPV66	-	Alpha	A	A1	U31794
HPV66	Qv25111	Alpha	B	B1	EF177188
HPV66	Qv25260	Alpha	B	B2	EF177187
HPV67	-	Alpha	A	A1	D21208
HPV67	Qv22701	Alpha	A	A2	HQ537780
HPV68	-	Alpha	A	A1	X67161
HPV68	Qv01017	Alpha	A	A2	KC470269
HPV68	Qv18016	Alpha	B	B1	KC470270
HPV68	C1	Alpha	C	C1	FR751039
HPV68	Qv33015	Alpha	C	C2	KC470274
HPV68	Qv00677	Alpha	D	D1	KC470275
HPV68	Qv30285	Alpha	D	D2	KC470276
HPV68	Qv17725	Alpha	E	E1	KC470277
HPV68	Qv25395	Alpha	F	F1	KC470279
HPV68	Rw826	Alpha	F	F2	KC470281
HPV69	-	Alpha	A	A1	AB027020
HPV69	Qv29573	Alpha	A	A2	KF436859
HPV69	Qv35103	Alpha	A	A3	KF436861
HPV69	Qv31811	Alpha	A	A4	KF436863
HPV70	-	Alpha	A	A1	U21941
HPV70	Qv5102	Alpha	B	B1	KC470287
HPV71	-	Alpha	-	-	NC_002644
HPV72	-	Alpha	-	-	X94164
HPV73	-	Alpha	A	A1	X94165
HPV73	QV23185	Alpha	A	A2	KF436829
HPV73	QV00749	Alpha	B	B1	KF436834
HPV74	-	Alpha	-	-	AF436130
HPV75	-	Beta	-	-	Y15173
HPV76	-	Beta	-	-	Y15174
HPV77	-	Alpha	-	-	Y15175
HPV78	-	Alpha	-	-	KC138720
HPV80	-	Beta	-	-	Y15176
HPV81	-	Alpha	-	-	AJ620209
HPV82	-	Alpha	A	A1	AB027021
HPV82	Qv34571	Alpha	A	A2	KF436787
HPV82	Qv28313	Alpha	A	A3	KF436793
HPV82	Rw17	Alpha	B	B1	KF436794
HPV82	P5_860	Alpha	B	B2	KF444055
HPV82	Rw606	Alpha	C	C2	KF436800
HPV82	Rw74	Alpha	C	C3	KF436801
HPV82	Rw15	Alpha	C	C4	KF436802
HPV82	Qv28248	Alpha	C	C5	KF436803
HPV83	-	Alpha	-	-	AF151983

HPV84	-	Alpha	-	-	AF293960
HPV85	-	Alpha	-	-	AF131950
HPV86	-	Alpha	-	-	AF349909
HPV87	-	Alpha	-	-	AJ400628
HPV88	-	Gamma	-	-	NC_010329
HPV89	-	Alpha	-	-	AF436128
HPV90	-	Alpha	-	-	NC_004104
HPV91	-	Alpha	-	-	AF419318
HPV92	-	Beta	-	-	NC_004500
HPV93	-	Beta	-	-	AY382778
HPV94	-	Alpha	-	-	AJ620211
HPV95	-	Gamma	-	-	AJ620210
HPV96	-	Beta	-	-	NC_005134
HPV97	-	Alpha	-	-	EF202168
HPV98	-	Beta	-	-	NC_012744
HPV99	-	Beta	-	-	NC_012745
HPV100	-	Beta	-	-	NC_012746
HPV101	-	Gamma	-	-	NC_008189
HPV102	-	Alpha	-	-	DQ080083
HPV103	-	Gamma	-	-	NC_008188
HPV104	-	Beta	-	-	NC_012750
HPV105	-	Beta	-	-	NC_012747
HPV106	-	Alpha	-	-	DQ080082
HPV107	-	Beta	-	-	EF422221
HPV108	-	Gamma	-	-	NC_012213
HPV109	-	Gamma	-	-	NC_012485
HPV110	-	Beta	-	-	EU410348
HPV111	-	Beta	-	-	EU410349
HPV112	-	Gamma	-	-	NC_012486
HPV113	-	Beta	-	-	NC_012748
HPV114	-	Alpha	-	-	NC_013931
HPV115	-	Beta	-	-	NC_013591
HPV116	-	Gamma	-	-	NC_013035
HPV117	-	Alpha	-	-	GQ246950
HPV118	-	Beta	-	-	GQ246951
HPV119	-	Gamma	-	-	GQ845441
HPV120	-	Beta	-	-	GQ845442
HPV121	-	Gamma	-	-	NC_014185
HPV122	-	Beta	-	-	GQ845444
HPV123	-	Gamma	-	-	GQ845445
HPV124	-	Beta	-	-	GQ845446
HPV125	-	Alpha	-	-	FN547152
HPV126	-	Gamma	-	-	NC_016157
HPV127	-	Gamma	-	-	NC_014462
HPV128	-	Gamma	-	-	NC_014952
HPV129	-	Gamma	-	-	NC_014953

HPV130	-	Gamma	-	-	GU117630
HPV131	-	Gamma	-	-	NC_014954
HPV132	-	Gamma	-	-	NC_014955
HPV133	-	Gamma	-	-	GU117633
HPV134	-	Gamma	-	-	NC_014956
HPV135	-	Gamma	-	-	NC_017993
HPV136	-	Gamma	-	-	NC_017994
HPV137	-	Gamma	-	-	NC_017995
HPV138	-	Gamma	-	-	HM999990
HPV139	-	Gamma	-	-	HM999991
HPV140	-	Gamma	-	-	NC_017996
HPV141	-	Gamma	-	-	HM999993
HPV142	-	Gamma	-	-	HM999994
HPV143	-	Beta	-	-	HM999995
HPV144	-	Gamma	-	-	NC_017997
HPV145	-	Beta	-	-	HM999997
HPV146	-	Gamma	-	-	HM999998
HPV147	-	Gamma	-	-	HM999999
HPV148	-	Gamma	-	-	NC_014835
HPV149	-	Gamma	-	-	GU117629
HPV150	-	Beta	-	-	FN677755
HPV151	-	Beta	-	-	FN677756
HPV152	-	Beta	-	-	JF304768
HPV153	-	Gamma	-	-	JN171845
HPV154	-	Gamma	-	-	NC_021483
HPV155	-	Gamma	-	-	JF906559
HPV156	-	Gamma	-	-	JX429973
HPV159	-	Beta	-	-	HE963025
HPV160	-	Alpha	-	-	AB745694
HPV161	-	Gamma	-	-	JX413109
HPV162	-	Gamma	-	-	JX413108
HPV163	-	Gamma	-	-	JX413107
HPV164	-	Gamma	-	-	JX413106
HPV165	-	Gamma	-	-	JX444072
HPV166	-	Gamma	-	-	NC_019023
HPV167	-	Gamma	-	-	NC_022892
HPV168	-	Gamma	-	-	KC862317
HPV169	-	Gamma	-	-	JX413105
HPV170	-	Gamma	-	-	JX413110
HPV171	-	Gamma	-	-	KF006398
HPV172	-	Gamma	-	-	KF006399
HPV173	-	Gamma	-	-	KF006400
HPV174	-	Beta	-	-	HF930491
HPV175	-	Gamma	-	-	KC108721
HPV178	-	Gamma	-	-	NC_023891
HPV179	-	Gamma	-	-	NC_022095

HPV180	-	Gamma	-	-	KC108722
HPV184	-	Gamma	-	-	HG530535
HPV197	-	Gamma	-	-	KM085343
HPV199	-	Gamma	-	-	KJ913662
HPV200	-	Gamma	-	-	KP692114
HPV201	-	Gamma	-	-	NC_027528
HPV202	-	Gamma	-	-	KP692116
HPV204	-	Mu	-	-	KP769769
